JET PROPULSION LABORATORY
CALIFORNIA INSTITUTE OF TECHNOLOGY
PASADENA, CALIFORNIA

JPL-EP-204

External Publication No. 204

THE DESIGN OF OPTIMUM LINEAR SYSTEMS

Eberhardt Rechtin

JET PROPULSION LABORATORY
California Institute of Technology
Pasadena, California
April 1953

126 pages

TABLE OF CONTENTS

Table of Contents, Cont'd.

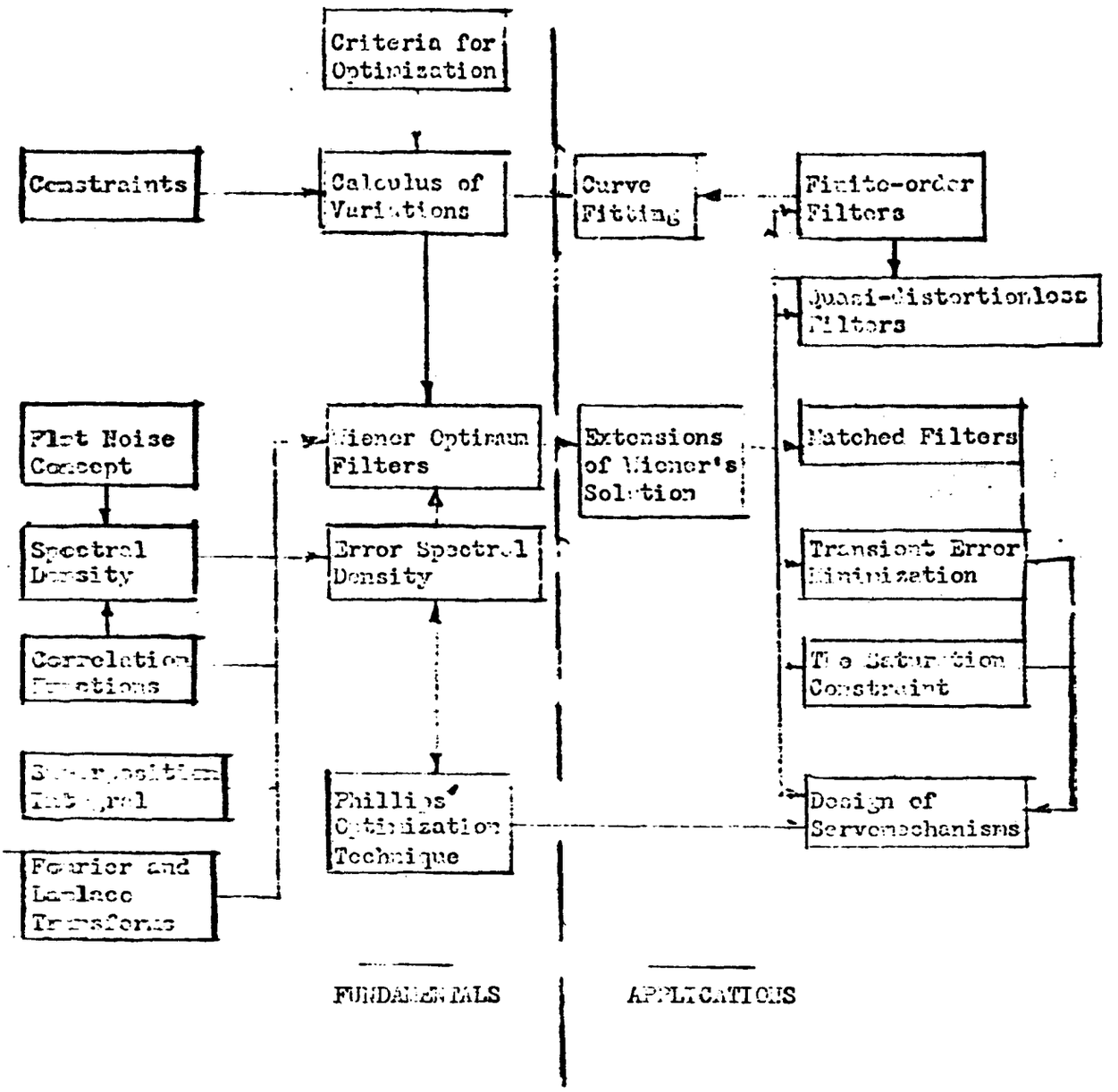


FIGURE 1
The Structure of Optimization Theory

INTRODUCTION

It is the purpose of this text to instruct engineers in the design of optimum linear systems. An optimum linear system is defined as an unbiased, linear system which performs a desired linear operation with the minimum mean square error. The basic cause for the system error is usually an external disturbance of a simple statistical nature. System error need not be caused by a disturbance, however; the design technique is also applicable to transient design of systems with limited power capabilities.

The presentation of the technique is directed toward the design and analysis of mobile-vehicle control systems, systems in which the control unit and the element controlled (the vehicle) are independent except for radio or optical links. Noise, disturbances of any kind, and deliberate enemy jamming may enter the system easily. The engineer's job is first, the design of the best control system under the circumstances, and second, the analysis of performance data to determine the degree of success achieved. In a very real sense, this particular problem is the sum of many problems including filtering signals from noise, designing stable control systems, preventing internal saturation of a servomechanism, and analyzing noisy records. Figure 1 indicates the order and relationship of the topics which describe the new technique.

The technique and derivations are described in terms of Laplace transforms. The author has assumed the reader to have a working knowledge of Laplace transforms roughly equivalent to that attained by reading Gardner and Barnes, "Transients in Linear Systems." No knowledge of statistics is assumed. No statistics is presented in this development which cannot be learned in a few minutes.

Certain portions of this text must be understood thoroughly in order that the engineer be able to do more than plug in and grind out the fundamental equation:

$$H(s)_{\text{optimum}} = \frac{1}{\Psi(s)} \left(\frac{H_1(s)[\Phi_{ff} + \Phi_{fn}]}{\Psi(-s)} \right)_{\substack{\text{physically} \\ \text{realizable}}}$$

The significance of the equation will be readily appreciated as the development proceeds, but it is not always simple to formulate the problem at hand in such a way as to make the equation applicable. Regardless of background, the reader must understand Sections I - X before reading further sections. Sections XI, XII, and XIII present the minimum theory necessary to solve problems. Section XIV outlines the remainder of the text in order that the reader may decide which further sections are directly related to his immediate problem.

The existence of the technique is directly attributable to Norbert Wiener, Professor of Mathematics at M.I.T. A. N. Kolmogoroff derived somewhat similar results but in a less useful form. With few exceptions, all extensions of Wiener's work were done in many places by many different people, all at about the same time. The derivations in this text are original in the sense of being independently derived at J.P.L. This text could well have been written by several other engineers at J.P.L. W. H. Pickering, F. W. Lehan, and R. J. Parks initiated and guided the work. C. W. Bergman, R. B. Conn, E. Rechtin, W. F. Sampson, R. M. Stewart, and D. C. Youla continued the development. This text is one result.

I. CRITERIA FOR OPTIMIZATION

Some amazingly violent arguments can ensue on the subject of optimization. Because of the great variety of possible systems, all performing the same task, it is difficult to get agreement even on the relative values of cost, reliability, simplicity, and accuracy. This discussion will be limited to considerations of accuracy for an extremely important reason -- when interference is present, it is not possible to make a system as accurate as we please. We (possibly) can increase cost without limit, and can make a system as reliable as desired, but we are definitely limited in the attainable accuracy for the system. The limiting accuracy of the system depends primarily upon the characteristics of the input functions -- loosely speaking, upon the characteristics of signal and noise -- and is realized only by the optimum system. If this limited accuracy is insufficient for our purposes, we must either change the input characteristics or abandon the attempt. No ammount of cost, complexity, or ingenuity can yield any improvement.

The existence of an optimally-accurate system and a knowledge of its associated "irreducible error" is useful even though such a system may occasionally be too complicated, costly, and/or unreliable to build. Such a system provides a good standard for evaluation of proposed substitutes. It is often possible to design substitutes with accuracies within a few percent of the limiting accuracy.

Let us define an optimum system as optimum in an accuracy sense. Again, there is much room for differences of opinion. For example, if the problem is to transmit information in the presence of noise, systems of different

relative accuracies may be designed depending on the coding (AM, FM, PCM, random) and upon the channel noise level. If the information may be quantized, the number of practical codes may be still further increased. However, for each transmitted code and assumed interference, there still exists an optimum receiver design. The problem becomes even more interesting if the interference is assumed intelligent, i.e., if the interference is always of the worst possible type with respect to the selected code. This text will not consider such problems in the theory of games, but will be restricted to optimum systems with respect to specified inputs.

If all inputs to a system are completely specified functions of time except for certain parameters such as absolute magnitude or exact time of occurrence, elimination of the effects of any of the individual inputs may usually be accomplished identically. 60 cycle "hum-bucking" is one example of complete interference elimination. Output hum consists of a pure 60 cycle tone which is eliminated by addition of an equal and opposite 60 cycle tone to the output. As another illustration, if the inputs are exactly known except for three constant parameters, three operations are usually sufficient to yield error-free performance. If the inputs are statistical in nature, however, there are so many unknowns that exact separations are impossible. We can only hope for good performance averaged over many tries.

It is important that we know beforehand in what way the various inputs differ from each other. If all inputs enter the system at the same point and all are alike, separation is impossible. Conversely, the greater the differences, the better the separation. Inputs may be described by their exact time dependence, by their expected time dependence, by their complete probability functions, or by their correlation functions (or spectral densities).

Statistical descriptions are comparatively weak, but they may be the only descriptions available. To quote Wiener, "Statistical prediction is essentially a method of refining a prediction which would be perfect by itself in an idealized case but which is corrupted by statistical errors, either in the observed quantity itself or in the observation. Geometrical facts must be predicted geometrically and analytical facts analytically, leaving only statistical facts to be predicted statistically."

We complete the specification of an optimum system by agreeing on a mathematical description of system accuracy. If no restrictions are placed upon either the probability functions of the inputs, or on the type of operations performed, it is possible to specify the accuracy of the system in many different ways. For example, using the "maximum likelihood" criterion, one attempts to form the observed input by the addition of samples selected from known distributions in such a way that the joint probability of occurrence of the samples is maximized. The selection process may well be non-linear. A linear operation is describable by an equation of the first degree in the dependent variable. Such equations may be time-variant, integro-differential or difference, but may not involve operations on other than the first power of the dependent variable. Accuracy specifications involving probability functions generally lead not to explicit specification of the operation to be performed, but only to certain criteria on such operations. The field of non-linear mathematics is not yet in a condition to be extensively exploited by engineers. Each non-linear problem is approached independently. No general discipline is available at this date.

If we limit ourselves to linear systems,* not only do we enter a well-developed field, but we can also define system accuracy in a simple way. Let

$x(t)$ = desired output at time t

$x_o(t)$ = actual output at time t

$\varepsilon(t) = x(t) - x_o(t)$ = error at time t

Let us repeat the experiment many times and observe the performance at time t. The performance will be different for each experiment since the input interferences will never be exactly alike. Not knowing beforehand the exact time dependence of the interference, we can never guarantee perfect performance (zero error) for any one experiment. After observing many experiments, however, we certainly want to notice good overall performance. We desire no error on the average. Letting a bar signify experimental (ensemble) average, we require the condition:

$$\overline{\varepsilon(t)} = 0$$

Systems satisfying this condition are called "unbiased." The condition is not a good measure of comparative performance of various systems. Almost any system will yield $\overline{\varepsilon(t)} = 0$ inasmuch as the average interference is usually zero. Let us keep this condition as a desired feature and investigate several additional measures of system accuracy. Inasmuch as negative errors are probably as serious as positive errors, the measure should so indicate.

---

*The restriction to linear systems is not as stringent as it might seem. Disturbances produced by most interference phenomena are characterized by Gaussian distribution functions. Any linear operation on such disturbances will yield an output which also has a Gaussian distribution function. If we restrict our desired operations to linear operations (i.e., do not look for the square of the input, etc.), it has been shown that for Gaussian type input functions the optimum linear system is also the optimum of all systems. In addition, all standard descriptions of accuracy reduce to the one presented here.

Large errors are probably more serious than small errors.* Two measures of accuracy are of particular interest:

$$\overline{\varepsilon^2} \quad \text{and} \quad \overline{|\varepsilon|}$$

Of these, $\overline{\varepsilon^2}$ (the mean square or variance of ε), is in wide use, has been extensively developed, and leads to linear systems. The second, $\overline{|\varepsilon|}$, is extremely difficult mathematically, probably yields performance similar to $\overline{\varepsilon^2}$, and yet probably requires non-linear operations. Thus our definition of an optimum system: AN OPTIMUM LINEAR SYSTEM IS DEFINED AS AN UNBIASED, LINEAR SYSTEM WHICH PERFORMS A DESIRED LINEAR OPERATION WITH THE MINIMUM MEAN SQUARE ERROR.

---

*It might be agreed, however, that system design should ignore a fantastically large error on the basis that if such an error does occur, that particular experiment is worthless anyway, and should not be allowed to influence our decision on an optimum system. Such an argument serves to reject criteria such as $\overline{\varepsilon^4}$, $\overline{\varepsilon^6}$ etc. A system operating with an $\overline{\varepsilon^4}$ criterion would tend to average noise peaks and disregard errors close to zero.

II ELEMENTARY CALCULUS OF VARIATIONS

The optimum system has been defined as that system which minimizes a particular quantity, the mean square error. If our system is completely specified except for one parameter such as the value of a particular resistor or the gain of a particular servo loop, the solution for an optimum system is quite simple. We express the error σ^2 as a function of the variable parameter (r) and use the usual calculus technique

$$\frac{d\sigma^2}{dr} = 0 \qquad (1)$$

to find the minimum. This equation specifies a stationary point only (maximum, minimum, or inflection point) and therefore a check is usually made of the sign of $d^2\sigma^2/dr^2$. A positive sign indicates a true minimum. If the system is specified except for possibly four or five parameters, the general procedure is the same. True minimum can be realized providing the four or five simultaneous equations

$$\frac{d\sigma^2}{dr_1} = 0 \quad \frac{d\sigma^2}{dr_2} = 0 \quad \cdots \quad \frac{d\sigma^2}{dr_5} = 0 \qquad (2)$$

can be solved properly. It is not generally true, however, that by making more and more parameters variable we come nearer and nearer to the optimum system. The very act of writing the error as a function of certain parameters also defines the form of the system. For example, calculation of the effect of varying the value of a given resistor depends upon knowledge of the rest of the circuit. In particular, there is no way of knowing which of the circuits below will yield the best performance without laborious calculation. More important, there is no way of knowing whether the best circuit is even included

in this set.

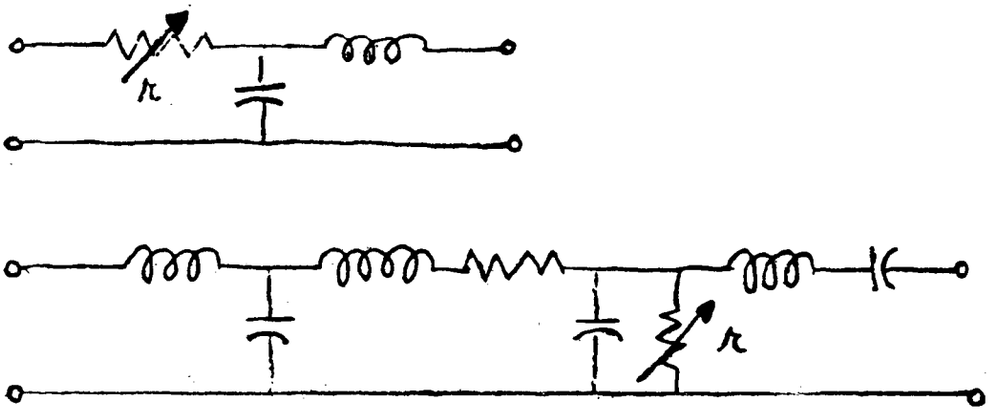


FIGURE 2
Examples of Circuits with an Adjustable Element

Even making all the elements variable will serve no useful purpose: There is no guarantee that the correct type element is located in each branch.

It is probably not surprising that a solution exists which not only gives the form of the best circuit, but also the values of all the elements in it. It is almost the engineer's creed that nothing is impossible. It may be surprising (1) that the concepts and techniques are quite simple and (2) that most of these concepts and techniques are 250 years old.

The magic mathematics is the calculus of variations. Its prime uses it seems, have been to prove that the shortest distance between two points is a straight line, or to derive equations of motion of a system in peculiar coordinate systems. For our purpoces both such problems are unnecessarily complicated. The basic problem in the calculus of variations is the minimization of an integral whose integrand is some function of the system. As we shall see later, the mean square error of a system is expressible as an integral whose

integrand depends on the system transfer function. For the present, the exact way in which we obtain some integrals as are presented below is not important. Of immediate importance is the question: what should H(s) be as a function of s to minimize the value of the integral?*

$$I_1 = \int_0^1 \left[s^2 H_1(s) + \frac{s^4}{H_1(s)} \right] ds$$

$$I_2 = \int_0^{+j\infty} \left\{ |H_2(s)|^2 \cos s + [1 - H_2(s)]^2 e^{-s} \right\} ds \qquad (3)$$

$$I_3 = \int_{-j\infty}^{+j\infty} \left\{ \Phi_1(s)[H_3(s)]^2 + \Phi_2(s)[1 - H_3(s)]^2 \right\} ds$$

where $\Phi_1(s)$ and $\Phi_2(s)$ are known.

The basic logic of the calculus of variations is as follows:

1. Assume that the correct answer is $H_o(s)$.
2. Pick out any other function of s, $\eta(s)$.
3. Add $\eta(s)$ to $H_o(s)$, but control the addition by a numerical variable ε. Th variable ε is not a function of s. It is the <u>only</u> variable, however, since both $H_o(s)$ and $\eta(s)$ are

---

* Answers:

$$H_1(s) = s \qquad H_2(s) = \frac{e^{-s}}{e^{-s} + \cos s} \qquad H_3(s) = \frac{\Phi_2(s)}{\Phi_1(s) + \Phi_2(s)}$$

presumed known. Then write H(s) near $H_o(s)$ as:

$$H(s) = H_o(s) + \varepsilon \eta(s)$$

4. Investigate the effect of changing ε on the integral of interest. Since we have assumed that we know $H_o(s)$ and $\eta(s)$ we could, formally, at least, plot the integral (I) as a function of ε.

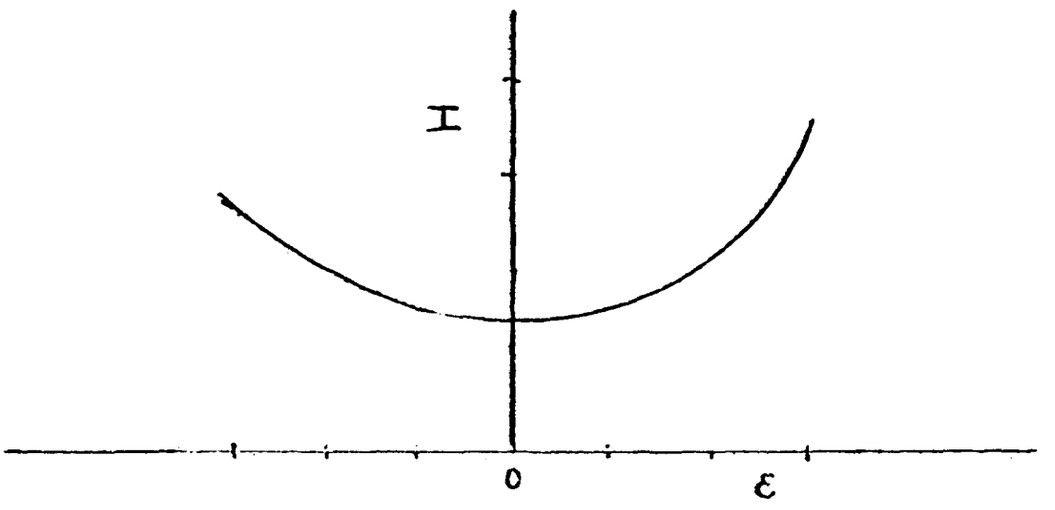


FIGURE 3
Value of the Minimized Integral as a Function of the Variation ε

By hypothesis I for $\varepsilon=0$ is the minimum I since H(s) for $\varepsilon=0$ is equal to the right answer $H_o(s)$.

5. Pick out many $\eta(s)$ and plot the resulting I's.

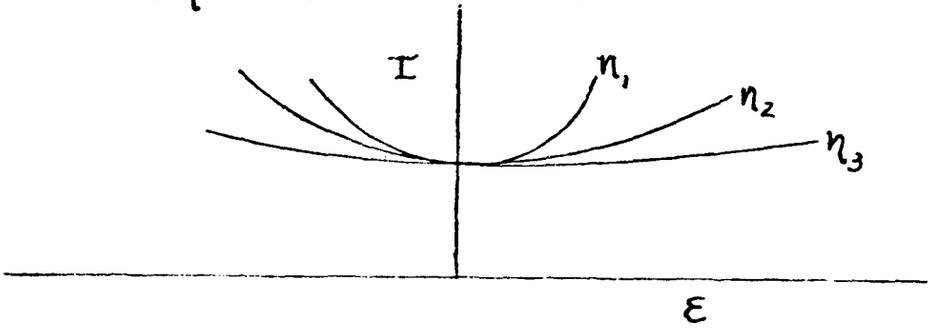


FIGURE 4
Variation of the Minimized Integral as a Function of ε and $\eta(s)$

It now becomes apparent that if $\eta(s)$ is not too unreasonable the choice of $\eta(s)$ does not affect the fact that at $\varepsilon = 0$ we have a minimum*, i.e., that independent of $\eta(s)$

$$\left.\frac{dI}{d\varepsilon}\right|_{\varepsilon=0} = 0 \qquad (4)$$

Based on this logic, the technique for solving for the H(s) which minimizes the integral I is:

1. Substitute $[H_o(s) + \varepsilon\eta(s)]$ for H(s) in the integral
2. Find
$$\left.\frac{d\,I(\varepsilon)}{d\,\varepsilon}\right|_{\varepsilon=0} = 0$$
3. Make this equation true regardless of $\eta(s)$
4. If necessary, check the sign of the second derivative $d^2I/d\varepsilon^2$

Example:

$$I = \int_0^1 \left[s^2 H(s) + s^4/H(s) \right] ds$$

$$I(\varepsilon) = \int_0^1 \left\{ s^2[H_o(s) + \varepsilon\eta(s)] + \frac{s^4}{H_o(s) + \varepsilon\eta(s)} \right\} ds$$

$$\frac{dI(\varepsilon)}{d\varepsilon} = \int_0^1 \left\{ s^2\eta(s) - \frac{s^4\,\eta(s)}{[H_o(s) + \varepsilon\eta(s)]^2} \right\} ds$$

$$\left.\frac{d\,I(\varepsilon)}{d\,\varepsilon}\right|_{\varepsilon=0} = \int_0^1 \eta(s)\left[s^2 - \frac{s^4}{H_o(s)^2} \right] ds$$

which is zero independent of $\eta(s)$ only if

$$H_o(s) = \pm s$$

*For stationary point. The condition for a minimum is a positive sign of $d^2I/d\varepsilon^2$ at $\varepsilon = 0$, independent of the choice of $\eta(s)$.

The positive sign yields a minimum value for the integral.

Problem 1 a: Find the function H(s) which minimizes I_3 in the text.

Problem 1 b: Find the function H(s) which minimizes

$$I = \int_A^B \left\{ 1 + \left[\frac{d\,H(s)}{d\,s} \right]^2 \right\} d\,s$$

Problem 2: It is also possible to solve normal calculus minimization problems by the variational method. The variational method used in this way is somewhat like using a cannon to kill a moth, but the use is still legitimate. The only difference in logic is that η is an arbitrary value rather than an arbitrary function. Solve for a minimum of

$$y = x^2 - 2x$$

using the variational technique.

For the mathematical philosophers among the readers, the integral is not the only mathematical operation to which this technique is applicable. Any operation is valid which reduces the performance of a system to a single number. A summation of discrete values is thus legitimate, as are such parameters as joint and marginal distributions in probability problems.

III CONSTRAINTS

A constraint is a restriction or condition imposed on the solution of a problem. Constraints in optimization theory may either reduce system performance or increase it, depending on the particular constraint. Constraints such as

1. The available power is limited at certain points within the system,
2. The system must be physically realizable, and
3. Certain functions of time (polynomials or sine waves) must pass through the system without distortion

will all change system performance from that of the unconstrained system.

Constraints are encountered on all levels of calculus.* A typical calculus problem involving a constraint is:

What is the minimum value of

$$y = \sqrt{1 - (x-1)^2 - z^2}$$

under the condition

$$x + y + z = 1 \ ?$$

Of more direct interest to us is the analogous problem in integral form:

What is the minimum value of

$$I_1 = \int_0^1 \left[s^2 H(s) + \frac{s^4}{H(s)} \right] ds$$

under the constraint

$$I_2 = \int_0^1 \left[H(s) \right] ds = 2.0$$

---

* Sokolnikoff, Advanced Calculus, page 327

These two problems involve single constraints only. Problems may be proposed with any number of constraints. For example, we might add the constraint

$$I_3 = \int_0^1 \frac{s^2}{H(s)}\, ds = 1.2$$

on to the second illustrative problem. The constraints all possess the property that, being constants, their variation with ϵ is zero. In the above example, if the substitution $H_0(s) + \epsilon \eta(s)$ were made in I_2 and I_3

$$\left.\frac{d I_2(\epsilon)}{d \epsilon}\right|_{\epsilon = 0} = \frac{d}{d\epsilon}(2.0) = 0$$

$$\left.\frac{d I_3(\epsilon)}{d \epsilon}\right|_{\epsilon = 0} = \frac{d}{d\epsilon}(1.2) = 0$$

Thus, if λ_1 and λ_2 (the so-called Lagrange multipliers) are constants, it is certainly true that

$$\left[\frac{d I_1}{d \epsilon} + \lambda_1 \frac{d I_2}{d \epsilon} + \lambda_2 \frac{dI_3}{d\epsilon}\right]_{\epsilon = 0} = 0 \qquad (5)$$

The above equation is a special way of substituting I_2 and I_3 into I_1 before differentiation. It can be shown that

$$\frac{d}{d \epsilon}\left(I_1 + \lambda_1 I_2 + \lambda_2 I_3\right)\Big|_{\epsilon = 0} = 0$$

is both necessary and sufficient to specify a stationary value of I_1 under the constraints I_2 and I_3. In this particular problem, therefore, the procedure is to minimize

$$I_1 + \lambda_2 I_2 + \lambda_3 I_3 = \int_0^1 \left[s^2 H + \frac{s^4}{H} + \lambda_1 H + \frac{\lambda_2 s^2}{H} \right] ds$$

in the same way that I_1 was minimized alone. The result of the minimization is the equation

$$H(s) = \pm s\sqrt{\frac{s^2 + \lambda_2}{s^2 + \lambda_1}}$$

The constants λ_1 and λ_2 are evaluated from the equations $I_2 = 2.0$ and $I_3 = 1.2$.

Notice that the minimization procedure is symmetrical in I_1, I_2, and I_3. There is no difference, therefore, between minimizing I_1 with I_2 and I_3 fixed, between minimizing I_2 with I_1 and I_3 fixed, and between minimizing I_3 with I_1 and I_2 fixed. This symmetry is occasionally useful in re-stating a problem; symmetry will be used in section XIX in precisely this way.

Problem 3 : Find the H(s) which minimizes

$$I_1 = \int_0^1 (1 - s^2)\, H(s)\, ds$$

under the constraint

$$I_2 = \int_0^1 \frac{ds}{H(s)} = 1.0$$

$\min(I_1 + \lambda I_2) = \min I$

$I = \int_0^1 \left[(1-s^2)H(s) + \frac{\lambda}{H(s)}\right] ds$

Let $H(s) = H^*(s) + \epsilon\,\eta(s)$ and take $\frac{\partial I}{\partial \epsilon}\Big|_{\epsilon=0} = 0$

$0 = \frac{\partial I}{\partial \epsilon}\Big|_{\epsilon=0} = \int_0^1 \left[(1-s^2)\,\eta(s) - \frac{\lambda\,\eta(s)}{[H^*(s)]^2}\right] ds$

Set $\frac{\partial I}{\partial \epsilon}\Big|_{\epsilon=0} = 0$ independ of the $\eta(s)$ chosen

Problem 4: Find the h(t) which minimizes

$$\sigma^2 = \int_{-T}^{+T} h^2(t)\, dt$$

under the constraints

$$1 = \int_{-T}^{+T} h(t)\, dt \quad \text{and} \quad 0 = \int_{-T}^{+T} t\, h(t)\, dt$$

$\therefore [H^*(s)]^2 = \frac{\lambda}{[1-s^2]}$

min and λ simply set

$1 = I_2 = \int_0^1 \frac{1}{\sqrt{\lambda}} \sqrt{(1-s^2)}\, ds$

$\sqrt{\lambda} = \frac{\pi}{4}$

$H^*(s) = \frac{\pi}{4}\sqrt{1-s^2}$

IV. CURVE FITTING AND WEIGHTING FUNCTIONS

The idea of optimizing system performance is almost as old as the calculus itself. One of the first problems in statistics is that of fitting the best analytical curve to a set of data points. This problem is worth a detailed study because it leads logically to such concepts as weighting functions and matched filters. Let us solve the problem in the classical manner first. Later we shall solve the same problem using variational techniques under constraint. Most of the basic ideas of optimization theory will appear in the course of these solutions.

A. Classical Statement of the Curve Fitting Problem

Every optimization problem requires certain a priori assumptions. The assumptions in the classical curve fitting problem are:

1. A mechanism exists which is producing a perfect, disturbance-free function. (An oscillator is producing a perfect sine wave. An object is moving under no external forces.)
2. The form of this function is known, although certain parameters are unknown. (Let the function be $x(t)$ where t is the independent variable. Then some examples of functional form might be

$$x(t) = a + b\,t$$

$$x(t) = a \sin (t + b)$$

where the parameters a and b are to be determined from the data.)
3. Our discrete observations of $x(t)$ are perturbed by a disturbance of statistical nature. In particular the disturbance has zero average value and a σ_N^2 mean square value. The statistical characteristics of the disturbance remain the same throughout the interval considered.

4. The criterion for optimization is the minimization of

$$\sigma^2 = \sum_{k=1}^{K} \left[x(k)_{estimate} - x(k)_{data} \right]^2 \qquad (6)$$

where K is the total number of observation points, $x(k)_{estimate}$ is the functional form with appropriately chosen parameters, and $x(k)_{data}$ represents the observed values.

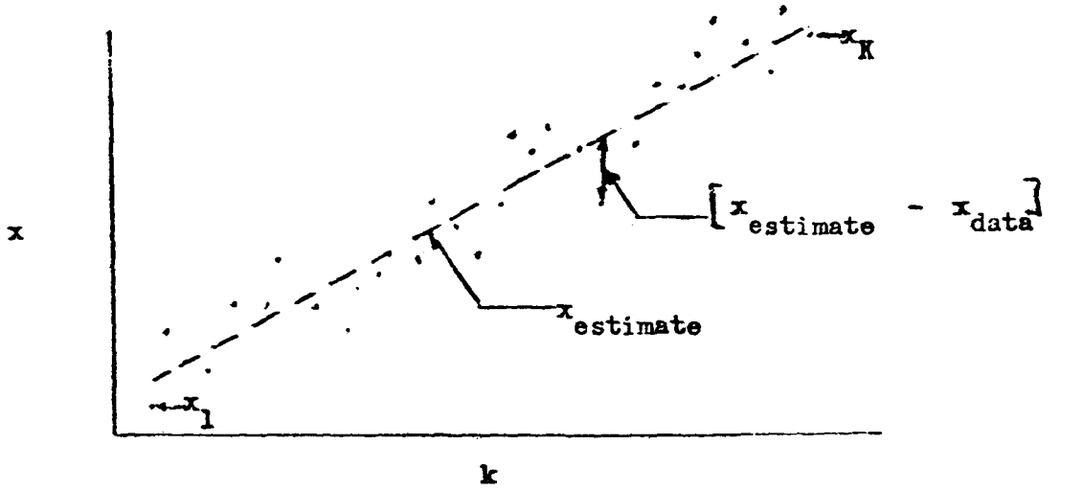


FIGURE 5
Typical Data Plot

The standard presentation of this curve fitting problem usually mentions a fifth assumption but seldom adds an explanation. The fifth assumption: the disturbance is such that all observations of it are independent. This assumption makes the overall field of discrete statistics considerably simpler, but immediately precludes the extension to the continuous case. In the continuous case, the samples are infinitely close together. A statement that adjacent samples are still independent under these conditions is not only

unbelievable, it also leads to mathematical explosions. The most important use of the independence assumption is in the proof that the optimization criterion of assumption 4 yields the best possible discrete estimator in the ensemble mean square sense. Let us deliberately ignore the independence assumption for the moment and note the consequences later.

The problem in curve fitting can thus be stated as follows. Under the four given assumptions, what is the best $x_{estimate}$ such that we minimize

$$\sigma^2 = \sum_{k=1}^{K} \left[x(k)_{estimate} - x(k)_{data} \right]^2$$

B. Classical Solution of the Curve Fitting Problem

For purposes of illustration, let us assume that the data points are equally spaced and that the assumed functional form is

$$x(k) = a + b\,k$$

The estimated value of x(k) is given by x^*

$$x(k)^* = a^* + b^* k$$

The mean square error σ^2 is given by

$$\sigma^2 = \sum_{k=1}^{K} \left[a^* + b^* k - x(k)_{data} \right]^2$$

The minimum value of which is given by

$$\frac{\partial \sigma^2}{\partial a^*} = 0 \qquad \text{and} \qquad \frac{\partial \sigma^2}{\partial b^*} = 0$$

These operations yield

$$0 = \sum_{k=1}^{K} \left[a^* + b^* k - x(k)_{data} \right]$$

and

$$0 = \sum_{k=1}^{K} \left[k \left(a^* + b^* k - x(k)_{data} \right) \right]$$

Solving for the parameters a^* and b^*

$$a^* = \frac{-\sum k^2 \sum x_{data} + \sum k \sum k x_{data}}{-K \sum k^2 + (\sum k)^2}$$

$$b^* = \frac{+\sum k \sum x_{data} - K \sum k\, x_{data}}{-K \sum k^2 + (\sum k)^2}$$

C. Extension of the Classical Solution

The above equations specifying the estimated values of the parameters are so complex in appearance that an extremely important concept is often missed. Noting that K, $\sum k$ and $\sum k^2$ are all constants, we may re-write the equations as:

$$a^* = + \sum_{k=1}^{K} \left[A_a - B_a k \right] x(k)_{data}$$

$$b^* = - \sum_{k=1}^{K} \left[A_b - B_b k \right] x(k)_{data} \qquad (7)$$

where A_a, B_a, A_b and B_b are constants dependent only on K. The process of estimating the parameters thus consists in weighting each data point according to an appropriate weighting function $\left[A - B\,k\right]$ and summing over all data points. It is interesting to superimpose the weighting functions on the data.

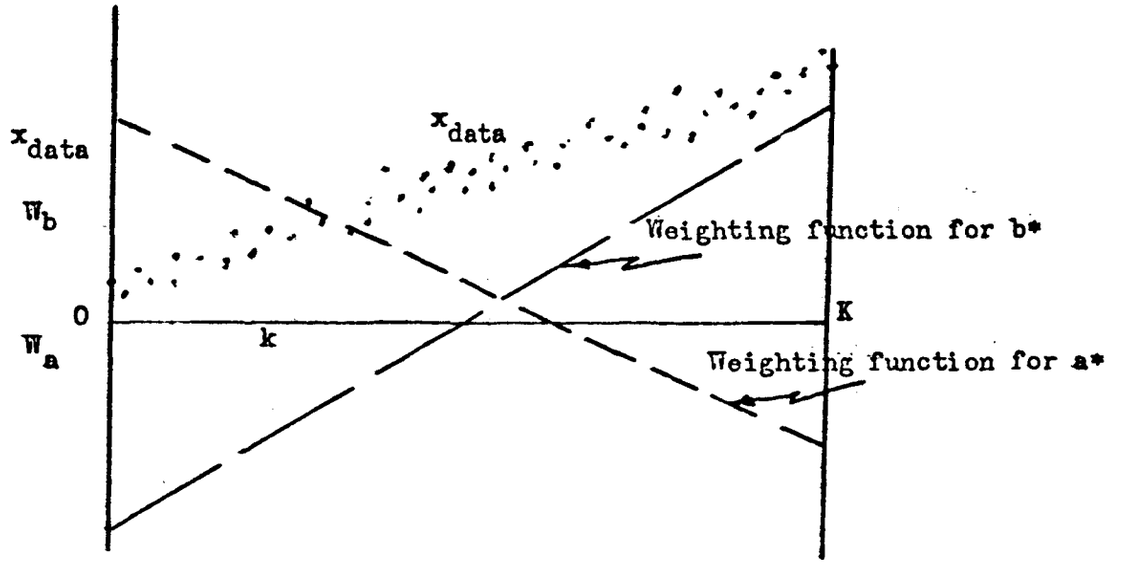


FIGURE 6
Weighting Functions

The similarity in form between the weighting functions and the functional form of x(k) occurs whenever the assumed form of x(k) is a sum of functions with unknown coefficients (and the noise samples are independent).

Problem 5: Assume a form $x(k) = a\, f_1(k) + b\, f_2(k)$, where $f_1(k)$ and $f_2(k)$ are known functions of k. Show that, as far as this type of optimization is concerned, the weighting functions for a* and b* are of a similar functional form.

Problem 6: A text in statistics claims that the problem of estimating a and b in the equation $x(k) = a\, e^{-bk}$ can be accomplished by taking logarithms of both sides of the equation and treating the system as linear. Show that this process does not minimize $\sum (x_{estimate} - x_{data})^2$ as was required for the best estimate.

This similarity in form of the weighting function to the assumed data function will appear frequently in later discussions. The similarity is again evident

in matched filter networks, finite-order filter networks, and cross-correlation detection networks.

D. Mean Square Error of the Estimated Parameters

If the disturbance producing the data scatter is strong, we should expect our estimates of parameters to be much less accurate than if the disturbance is weak. In particular, if σ_n^2 is the mean square value of the data scatter, W_b is the weighting function for the parameter b, and the data scatter samples are independent, then the mean square error on b is given by:

$$\sigma_b^2 = \sigma_n^2 \sum_1^K \left[W_b(k) \right]^2 \tag{8}$$

Proof:

Assume no signal $x(k)$ present, only the noise disturbance x_n. The estimate for the parameter b under these noise conditions is

$$b_n = \sum_1^K W_b(k)\, x_n(k)$$

If this experiment is performed many times, the average b_n will be zero provided the average $x_n(k)$ is zero. The mean square value of b_n is

$$\overline{b_n^2} = \overline{\sum_{k=1}^K W_b(k)\, x_n(k) \sum_{j=1}^K W_b(j)\, x_n(j)}$$

The right hand side of the equation is deliberately written as the product of two sums in different variables to emphasize the order in which the operations are performed. The trick is a common one and is useful in re-writing the mean square value of b_n as

$$\overline{b_n^2} = \sum_{k=1}^K W_b(k) \left[\sum_{j=1}^K W_b(j)\, \overline{x_n(k)\, x_n(j)} \right] \tag{9}$$

The $x_n(k)$ may be moved into the j summation since it is a constant with respect to j. The bracket is performed first, yielding a function of k which is multiplied by $W_b(k)$ and summed in k. The averaging operation affects only the x_n's because the W_b's are the same for every experiment.

The average product $x_n(k)\, x_n(j)$ is called the auto-correlation function of x_n. This correlation function will be discussed in detail in later sections. For this proof, it is only necessary to note that such an average product for independent x_n's is zero unless $k = j$. Providing that the statistics do not change during the experiment or between experiments, the mean square data scatter will be the same for all points k. In other words,

$$\overline{x_n(k_1)}^2 = \overline{x_n(k_2)}^2 = \sigma_n^2 = \text{constant independent of choice of k.}$$

The mean square error in estimating the parameter b in the presence of independent noise samples is thus given by

$$\sigma_b^2 = \overline{b_n^2} = \sigma_n^2 \sum_{k=1}^{K} \left[W_b(k) \right]^2 \qquad (8)$$

Any equation of the form

$$b = \sum_{k=1}^{K} W_b(k)\, x(k) \qquad (10)$$

is linear in x(k). Conversely, all linear operations may be written in the form of equation 10. The error equation, σ_b^2, thus applies to any linear operation in the presence of independent noise samples of mean square value σ_n^2.

E. Solution of the Curve Fitting Problem by the Calculus of Variations

The classical solution to the curve fitting problem is an application of the standard calculus. Attempts at extending the classical technique, however, usually result in such cumbersome mathematics that the basic logic is obscured. The calculus of variations technique does much to remedy the situation. An understanding of the curve-fitting solution to follow will mean an easy understanding of the Wiener derivation and the various related constraint problems (finite-order filters, matched filters, power level constraints, and servomechanism optimization).

The assumptions in this variational solution are quite similar to the classical assumptions. For variety, let us choose a slightly more general form for the assumed x(k).

1. Assume x(k) to be of the form

$$x(k) = a\, f_1(k) + b\, f_2(k) \tag{11}$$

where estimates of a and b are desired.

2. Assume the disturbance to be statistical in nature with a mean square value of σ_n^2 and an average of zero. The statistics of the disturbance are assumed the same throughout the experiment.

3. Assume that the criterion for the best estimate is the minimization of the mean square error of the estimated parameters. A subtle difference exists between this assumption and the equivalent classical one. The classical criterion minimizes the mean square difference between the data and the estimate over the interval. The calculus of variations criterion used here minimizes the mean square difference between the estimate and the correct answer over an ensemble. As will be seen, these two criteria are equivalent only if the noise samples are independent. In this problem, since the form of x(k) is linear in a and b, the weighting function representation is valid -- i.e., the system is linear -- and hence:

$$\begin{aligned} a^* &= \sum W_a(k)\; x(k)_{data} \\ b^* &= \sum W_b(k)\; x(k)_{data} \end{aligned} \tag{12}$$

The expressions for mean square parameter error for <u>independent</u> noise samples are

$$\sigma_a^2 = \sigma_n^2 \sum [W_a(k)]^2 \qquad \sigma_b^2 = \sigma_n^2 \sum [W_b(k)]^2 \tag{13}$$

If the noise samples are not independent, the expressions are more complicated. (See equations 8 and 9)

4. We desire that in the presence of no disturbance the W_a will yield a and the W_b , b. Thus

$$\begin{aligned} a &= \sum W_a(k) \left[a\, f_1(k) + b\, f_2(k) \right] \\ b &= \sum W_b(k) \left[a\, f_1(k) + b\, f_2(k) \right] \end{aligned} \tag{14}$$

Inasmuch as a and b are not necessarily related:

$$\left.\begin{aligned} 1 &= \sum W_a(k)\ f_1(k) \\ 0 &= \sum W_a(k)\ f_2(k) \end{aligned}\right\} W_a \text{ constraints}$$

$$\left.\begin{aligned} 0 &= \sum W_b(k)\ f_1(k) \\ 1 &= \sum W_b(k)\ f_2(k) \end{aligned}\right\} W_b \text{ constraints} \tag{15}$$

These equations express mathematically the constraint that a noise-free system yield the correct answers.

The problem to be solved is: 1. What is $W_a(k)$ such that σ_a^2 is minimized under the W_a constraints? 2. What is $W_b(k)$ such that σ_b^2 is minimized under the W_b constraints?

Solution for W_a :

$$\sigma_a^2 = \sigma_n^2 \sum W_a(k)^2$$

$$\begin{aligned} 1 &= \sum W_a(k)\, f_1(k) \\ 0 &= \sum W_a(k)\, f_2(k) \end{aligned}$$

Therefore, minimize

$$E_a^2 = \sigma_a^2 + \lambda_1 \left[1 - \sum W_a f_1 \right] + \lambda_2 \left[0 - \sum W_a f_2 \right] \tag{16}$$

and solve for necessary λ's. The minimization procedure: substitute $W_a(k) + \varepsilon \eta(k)$ for $W_a(k)$, differentiate with respect to ε at $\varepsilon = 0$ and make the result zero independent of η.

$$\frac{d\ E_a^2(\varepsilon)}{d\ \varepsilon} = \frac{d}{d\varepsilon}\left[\sigma_n^2 \sum (W_a + \varepsilon\eta)^2 + \lambda_1 - \lambda_1 \sum (W_a + \varepsilon\eta) f_1 - \lambda_2 \sum (W_a + \varepsilon\eta) f_2\right]$$

$$\left.\frac{d\ E_a^2}{d\ \varepsilon}\right|_{\varepsilon=0} = 0 = 2\sigma_n^2 \sum W_a\eta - \lambda_1 \sum f_1\eta - \lambda_2 \sum f_2\eta$$

$$= \sum_1^K \eta\left[2\sigma_n^2 W_a(k) - \lambda_1 f_1(k) - \lambda_2 f_2(k)\right]$$

(17)

Therefore

$$W_a(k) = \left[\frac{\lambda_1}{2\sigma_n^2}\right] f_1(k) + \left[\frac{\lambda_2}{2\sigma_n^2}\right] f_2(k) \qquad (18)$$

The λ's are evaluated by substituting equation 18 into the W_a constraints in equation 15. The form of $W_a(k)$ is again similar to the assumed form of x(k) and is easily shown to be the same weighting function derivable by the classical technique. The solution for $W_b(k)$ is exactly the same except for subscripts.

It is not surprising that the calculus of variations technique yields the same answer as the standard calculus technique under the same assumptions. It should. The variational technique shows the approach to the non-independent problem, however. Instead of the error expression

$$\sigma_b^2 = \sigma_n^2 \sum \left[W_b(k)\right]^2 \qquad (8)$$

we must use

$$\overline{b_n^2} = \sum_{k=1}^K W_b(k) \sum_{j=1}^K W_b(j)\ \overline{x_n(k)\ x_n(j)} \qquad (9)$$

Use of the $\overline{b_n^2}$ expression in the optimization process will yield a $W_b(k)$ which produces the smallest mean square error. Inasmuch as the classical solution yields an answer derivable from σ_n^2 and not $\overline{b_n^2}$, it can not result in the best $W_b(k)$ for non-independent noise samples. This fact may also be demonstrated using more elegant statistical arguments. It thus becomes obligatory to use the calculus of variations technique in solving the continuous $x(t)$ problem where the independence assumption is not realistic.

Problem 7: What are the values of λ_1 and λ_2 in the derived expression for $W_a(k)$ in equation 18?

Problem 8: What average error and mean square error should be expected in finding b where $x(k)$ is assumed of the form $x(k) = a + b\,k$ and where the noise consists of independent samples of mean square value σ_n^2 and average value zero?

V. FOURIER AND LAPLACE TRANSFORMS

The two-sided Fourier transform is a useful tool in certain sections of optimization theory. This transform is closely related to, and more restrictive than, the more familiar one-sided Laplace transform* defined by

$$F(s) = \int_0^{\infty} f(t)\, e^{-st}\, dt \tag{19}$$

$$f(t) = \frac{1}{2\pi j} \int_{-j\infty + c}^{+j\infty + c} F(s)\, e^{st}\, ds \tag{20}$$

where f(t) is the time function to be transformed, s is the so-called complex frequency, and c is a real number (often zero) used to guarantee integral convergence. Using the substitution s = jw in the conventional definitions of the two-sided Fourier transform shows its similarity to the Laplace transform.

$$\mathfrak{F}(s) = \int_{-\infty}^{+\infty} f(t)\, e^{-st}\, dt \tag{21}$$

$$f(t) = \frac{1}{2\pi j} \int_{-j\infty}^{+j\infty} \mathfrak{F}(s)\, e^{st}\, ds \tag{22}$$

The absence of the convergence number, c restricts the Fourier transform for our purposes to functions which satisfy

---

*Gardner & Barnes, "Transients in Linear Systems." The transform is called one-sided because only the positive time region is considered. A two-sided transform considers behavior in both positive and negative time.

$$\int_0^{\infty} |f(t)|^2 \, dt < \infty$$

$$\int_0^{-\infty} |f(t)|^2 \, dt < \infty \qquad (23)$$

Thus functions such as $e^{|\alpha t|}$ and $\cos \beta t$ have no Fourier transform. Providing the restrictions are met on f(t), however, tables of Fourier transforms are easily created from tables of Laplace transforms. From the defining integrals, if $F_1(s)$ is the one-sided Laplace transform of f(t) from t = 0 to $t = +\infty$ and if $F_2(s)$ is the one-sidod Laplace transform of f(t) going from t = 0 to $t = -\infty$, then

$$\mathcal{F}(s) = F_1(s) + F_2(-s)$$

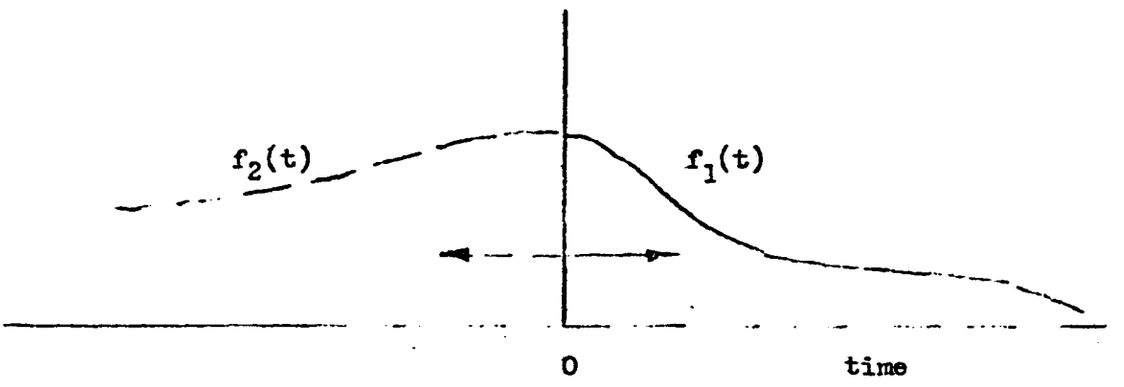


FIGURE 7

Time Functions

Problem 9: Find the two-sided Fourier transform of $e^{-\alpha|t|}$ and $e^{-\alpha|t|} \cos \beta t$

Problem 10: Prove that $\mathcal{F}(s) = F_1(s) + F_2(-s)$ as stated in the text.

The restriction on the Fourier transform provides an interesting and

useful means of determining whether f(t) occurred in +t or -t. This information is highly important in discussing the realizability of networks. For example: given

$$\mathcal{F}(s) = \frac{1}{\alpha - s} \qquad \alpha > 0$$

it might appear that two alternatives are possible.

1. $f_1(t)$ 0 for $t < 0$, $e^{\alpha t}$ for $t > 0$
2. $f_2(t)$ $e^{-\alpha |t|}$ for $t < 0$, 0 for $t > 0$

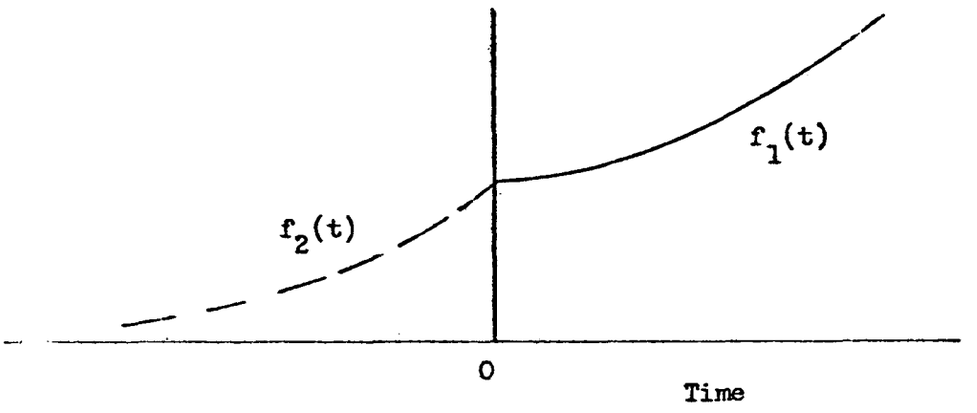


Figure 8

Alternative Time Functions

Only the second alternative satisfies the condition for existence of the Fourier transform. For the first alternative

$$\int_{-0}^{\infty} |f(t)|^2 \, dt = \int_{0}^{\infty} e^{2\alpha t} \, dt = \infty$$

and hence the Fourier transform does not exist. We might generalize this concept to show that $\mathcal{F}$-transforms of the type

$$\mathcal{F}(s) = \sum_{i=1} \frac{A_i}{\alpha_i + s} \qquad \text{Real } (\alpha_i) > 0$$

describe time functions in positive time only and that $\mathcal{F}$-transforms of the type

$$\mathcal{F}(s) = \sum_{i=1} \frac{A_i}{\alpha_i - s} \qquad \text{Real } (\alpha_i) > 0$$

describe negative-time functions.

The two-sided Fourier transforms possess another interesting property. Providing that the Wiener-Paley criterion

$$\frac{1}{2\pi j} \int_{-j\infty}^{+j\infty} \frac{|\log |\tilde{\alpha}_i(s)|\,|}{1 - s^2}\, ds < \infty \qquad (24)$$

is satisfied, the Fourier transform is said to be <u>factorable</u>: the factorable transform can be written as

$$\mathcal{F}(s) = \psi(s)\,\psi(-s) \qquad (25)$$

where $\psi(s)$ and $\psi(-s)$ are exactly alike except, (1) that $\psi(s)$ refers to a positive-time function and $\psi(-s)$ to a negative-time function and, (2) that $\psi(-s)$ can be written by substituting $(-s)$ everywhere for $(+s)$ in $\psi(s)$. As will be seen in later sections, the Wiener optimum solution depends upon being able to factor a Fourier transform into its two parts as defined in equation 25. The factorization theorem also has a practical significance with respect to realizable networks. Realizable networks have an impulsive response $h(t)$ of zero for $t < 0$ and consequently will have a transform $H(s)$ in positive s only. The amplitude response of the network is $|H(s)|$ and its attenuation $\log|H(s)|^2$. The criterion thus states that no realizable network can have infinite attenuation $\left[H(s_0) = 0\right]$

over any finite frequency band.

Example: Factor the $\mathcal{F}(s)$ for $f(t) = e^{-\alpha|t|}$

$$F_1(s) = \frac{1}{\alpha + s}$$

$$F_2(-s) = \frac{1}{\alpha - s}$$

$$\mathcal{F}(s) = \frac{2\alpha}{\alpha^2 - s^2}$$

which factors into

$$\mathcal{F}(s) = \frac{\sqrt{2\alpha}}{\alpha + s} \cdot \frac{\sqrt{2\alpha}}{\alpha - s}$$

Problem 11: Factor the $\mathcal{F}(s)$ for $f(t) = e^{-\alpha|t|} \cos \beta t$

Both the Laplace and the Fourier transforms are mathematical aids to simplify the solution of time-invariant systems. Such transforms are relatively worthless, however, in treating systems which vary with time.* A complete treatment of the optimization of linear systems certainly should include the time-variant case. Rather than develop an all-inclusive body of mathematics capable of solving both the variant and non-variant cases, however, we shall develop the non-variant solution only. There are good reasons for this approach. The only real difference between the all-inclusive theory and the time-variant theory is the level of mathematical labor. The basic solutions for optimum systems are identical in form. It is possible to carry the invariant case through the design stage, however, while the

---

* Typical systems are those in which the values of R's, L's, and C's vary with time.

variant case is blocked by a formidable integral equation which, in general, is insoluble.

Inasmuch as aerodynamicists and guidance engineers are often faced with time-variant systems, a rapid development should occur of the necessary mathematical aids to reduce the labor of time-variant system design. A reference of value is "Transforms for Linear Time-Varying Systems" by John A. Aseltine, a thesis submitted at U.C.L.A., Department of Engineering, in February 1952. The thesis is primarily a discussion of the use of other types of transforms than Laplace and Fourier. Bessel and Cauchy transforms are discussed in detail; methods of reducing variant systems to invariant systems by substitution are mentioned.

Problem 12: To show how time-variant equations may occasionally be reduced to invariant equations, reduce the following equation to a constant coefficient equation by substituting $u = e^t$

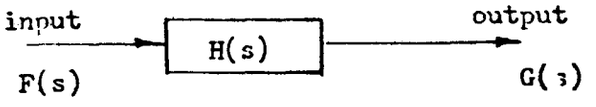
$$a u^2 \frac{d^2x}{du^2} + b u \frac{dx}{du} + c x = g(u)$$

In this equation x is the dependent variable; u is the independent variable; a, b, and c are constants; and $g(u)$ is a driving function.

VI THE SUPERPOSITION INTEGRAL

Electrical engineers have long been familiar with the use of transfer functions for description of network performance. In a linear invariant system the transfer function relates the output of the network to its input by means of appropriate Fourier or Laplace transforms. If F(s) represents the transform of the input, H(s) represents the transfer function, and G(s) represents the transform of the output, then

$$G(s) = H(s)\,F(s) \tag{25}$$



This same relationship may be written in the time domain* using the superposition integral.

$$g(t) = \int_0^t h(\tau)\, f(t-\tau)\, d\tau \tag{26}$$

where g, h, and f are the Laplace transforms of G, H, and F, respectively. The relationships are similar using two-sided Fourier transforms:

$$\mathcal{G}(s) = H(s)\,\mathcal{F}(s)$$
$$g(t) = \int_{-\infty}^{+\infty} h(\tau)\, f(t-\tau)\, d\tau \tag{27}$$

---

* Gardner and Barnes pages 228-236. Recommended reading.

Proof:

$$\mathcal{G}(s) = \int_{-\infty}^{+\infty} g(t)\, e^{-st}\, dt$$

$$= \int_{-\infty}^{+\infty} e^{-st} \left[\int_{-\infty}^{+\infty} h(\tau)\, f(t-\tau)\, d\tau \right] dt$$

$$= \int_{-\infty}^{+\infty} h(\tau) \left[\int_{-\infty}^{+\infty} f(t-\tau)\, e^{-st}\, dt \right] d\tau$$

$$= \int_{-\infty}^{+\infty} h(\tau)\, e^{-s\tau}\, dt \int_{-\infty}^{+\infty} f(t-\tau)\, e^{-s(t-\tau)}\, dt$$

$$= \int_{-\infty}^{+\infty} h(\tau)\, e^{-s\tau}\, d\tau \int_{-\infty}^{+\infty} f(u)\, e^{-su}\, du$$

$$= H(s)\, \mathcal{F}(s)$$

The time domain representation is more fundamental than a representation in the complex frequency domain. Time-variant linear systems, for example, are described by making the function h time-variable.

$$g(t) = \int_{-\infty}^{+\infty} h(\tau, t)\, f(t-\tau)\, d\tau \qquad (28)$$

The superposition integral is also useful as the clearest way of visualizing the cross-correlation operation used in detection of signals of known form. If the integral is re-written as a sum of discrete sampling operations, the resulting equation is recognizable as a weighting function equation similar to those given in the section on curve fitting.

$$g(K) = \sum_{k=1}^{K} h(k)\ f(K-k)$$

$$b = \sum_{k=1}^{K} W_b(k)\ x(k) \qquad \text{(equation 10, section IV C)}$$

The physical significance of the superposition integral may be visualized is several ways. If the input wave is plotted on a time scale as it arrives, then the network weighting function, $h(\tau)$, can be imagined as being dragged down the record at the same rate. The output is the integrated product of the two plots at each instant.

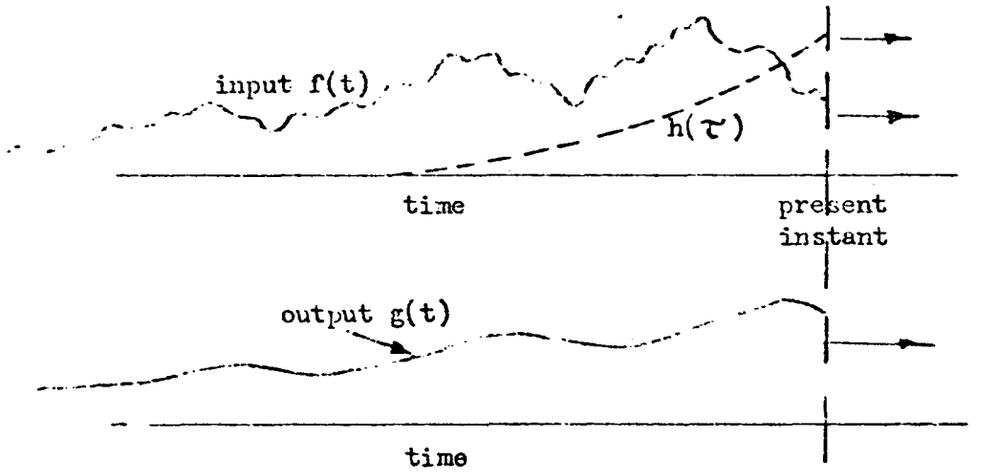


FIGURE 9
Graphical Picture of Superposition Integration

From the superposition integral it is evident that h(t) is the response of the network to an impulse. Thus the superposition integral might also be visualized as the(present) sum of responses to the(past)impulses defining f(t).

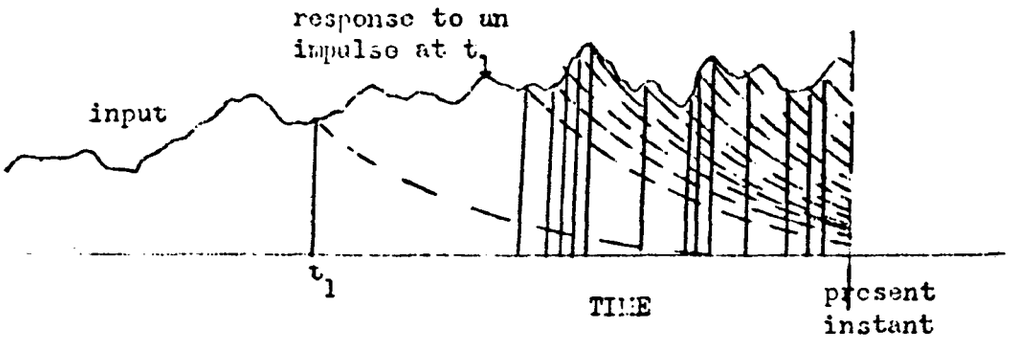


FIGURE 10
Graphical Picture of Superposition Integration

These visualizations illustrate the fact that the network $h(\tau)$ determines the present relative importance of the input signal τ seconds before. In this sense, $h(\tau)$ is a weighted memory.

The freedom with which we may choose $h(\tau)$ is limited in certain cases. In a physically-realizable system, it is impossible for the system to have a memory for events which have not yet happened. Thus, for physically-realizable networks:

$$h(-\tau) = 0 \qquad \text{where } \tau \text{ is a positive quantity} \tag{28}$$

This restriction is not always applicable. In data reduction work, all data with respect to the interval of interest is available, both past and future. Weighting functions extending in both directions in time are therefore admissible.

Problem 13: Sketch the response of a system described by the equations below to a step input at t = 0.

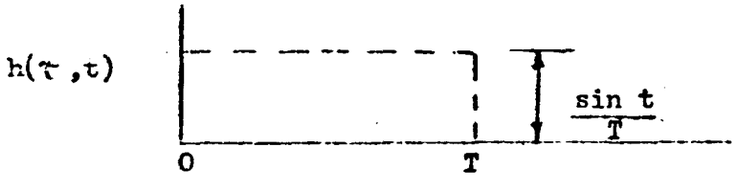
$$0 \leq \tau \leq T : \quad h(\tau) = 1/T$$

$$\tau > T : \quad h(\tau) = 0$$

Problem 14: Sketch the response to a time variant system described by the equations below to a step input at t = 0.

$$0 \leq \tau \leq T \quad : \quad h(\tau, t) = \frac{1}{T} \cdot \sin t$$

$$\tau > T \quad : \quad h(\tau, t) = 0$$



Problem 15: Sketch the weighting functions for the following:

a. An integrator
b. A pure time-delay
c. An imperfect integrator $H(s) = 1/(\alpha + s)$
d. A differentiator
e. An extremely narrow-band filter band-centered at f_o.

Problem 16: Demonstrate why the integral equation below cannot be Laplace transformed unless $f(t < 0) = 0$.

$$g(t) = \int_0^{+\infty} h(\tau)\, f(t - \tau)\, d\tau$$

Suggestion: Try it, using the same steps as given in the $\mathcal{L}$-transform proof of equation 27. Keep close track of the limits.

Note: This particular integral equation will appear later in section XII E with the conditions (1) that $f(t < 0) \neq 0$ and (2) the equation holds for $t > 0$ only. The solution of such an integral equation is fairly tricky.

VII CORRELATION FUNCTIONS

The concept of independent phenomena is easily described and understood. Two different random noise generators are undoubtedly independent. The static in radio reception is probably independent of the broadcast signal.

Related phenomena are also easily conceived, although the particular way in which relatedness is measured might seem arbitrary. The measure of relatedness is chosen to be the correlation function. Let f(t) be one function of time and g(t) another function of time. If these functions are compared in experiment after experiment, always at a time t_o after the experiment starts, we define the correlation function $\varphi_{fg}(t_o)$ to be the result of averaging $f(t_o)\ g(t_o)$ over many experiments.

$$\varphi_{fg}(t_o) = \overline{f(t_o)\ g(t_o)} \qquad (29)$$

The bar signifies the average over many experiments. If f(t) and g(t) repeat themselves identically in each experiment, the averaging process is unnecessary. On the other hand, if f(t) and g(t) are statistical in nature from experiment to experiment, the averaging process is of prime importance.

The major reason for measuring relatedness by the correlation function is that the correlation function (or its Fourier transform) is the one and only necessary function in deriving the optimum system.

Instead of comparing f(t) and g(t) both at t_o, we might compare f(t) at $t_o + \tau$ with g(t) at t_o. The correlation function then becomes

$$\varphi_{fg}(t_o, \tau) = \overline{f(t_o + \tau)\ g(t_o)}^{*} \qquad (30)$$

---

*The order of subscripts follows Wiener's convention. Phillips in M.I.T. Lab. 25 uses reverse order, gf.

This particular correlation function is called the cross-correlation function of f on g. Another comparison we might make is a comparison of $f(t_o + \tau)$ with $f(t_o)$. The result is the auto-correlation function of f.

$$\varphi_{ff}(t_o, \tau) = \overline{f(t_o + \tau)\, f(t_o)} \tag{31}$$

The correlation between a function and its value τ seconds away might well be expected to play an important part in system design. Systems have been mentioned as having weighted memories. A signal auto-correlation function decreasing slowly with τ indicates a slowly varying signal. Such signals are more easily extracted from noise than those with auto-correlation functions decreasing rapidly with time.* The correlation function measures the degree of relatedness. The auto-correlation function measures the amount that a signal hangs together in time and hence specifies how much of its past history is worth remembering.

The behavior of most statistical phenomena is such that the averaged performance is the same regardless of the particular instant t_o chosen for the observation. The behavior in time of such phenomena are called stationary time series. Notice that it is the average performance that is stationary. In stationary time series, the correlation functions do not depend upon t_o.

$$\varphi_{fg}(\tau) = \overline{f(t + \tau)\, g(t)}$$
$$\varphi_{ff}(\tau) = \overline{f(t + \tau)\, f(t)} \tag{32}$$

---

* Noise is usually characterized by auto-correlation functions decreasing rapidly with τ. The most difficult signals to extract from noise are those in which the signal auto-correlation closely resembles the noise auto-correlation.

Stationary time series are endowed with another property by hypothesis. The hypothesis states that it makes no difference whether the average is carried out at t_o for many experiments, or whether the average is carried out by using observations at different times in a single experiment. The ergodic hypothesis can not be proved, it can only be justified by experimental results. Some justification for the hypothesis might be found, however, by reasoning that because the time series is always present (we just aren't watching it) and because the time of start of the experiment is arbitrary, any observed value might well have occurred at the chosen t_o. Using the ergodic hypothesis and the integral definition of time average:

$$\varphi_{fg}(\tau) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{+T} f(t+\tau)\, g(t)\, dt \qquad (33)$$

In this development of optimization theory, the correlation function will be used primarily as a tool in derivations. Wiener has shown that the Fourier transform of the correlation function is the spectral density -- a somewhat easier starting point for the design engineer's imagination. The spectral density describes the power distribution of the function in the frequency domain. It is easier for the engineer to describe hum interference by remarking on the amount of 60 cycle power present than by stating that the correlation function exhibits a marked periodicity every 1/60 th of a second.

Problem 17: Find the auto-correlation function of $f(t) = a \sin(\beta t + C)$

There are two problems in which the correlation function itself is particularly useful. The first problem is the one posed by disturbances correlated with the signals. The second is the problem of signal detection by comparison of the incoming signal and noise with a locally generated signal. The latter problem is encountered in cross-correlation detection systems. When such systems can be used (detection of a sine wave in noise), they are capable of out performing conventional filter systems. The correlation functions are also used in reducing the labor required in empirical determination of spectral densities. For assistance in these problems, some of the properties of correlation functions are:

$$
\begin{aligned}
&1.\quad \varphi_{ff}(0) \geq \varphi_{ff}(\tau) \\
&2.\quad \varphi_{ff}(0) = \sigma_f^2 = \text{mean square value of } f(t) \\
&3.\quad \varphi_{ff}(\tau) = \varphi_{ff}(-\tau) \qquad (34) \\
&4.\quad \varphi_{fg}(\tau) = \varphi_{gf}(-\tau) \\
&5.\quad \varphi_{\dot{f}f}(\tau) = \frac{d}{d\tau}\varphi_{ff}(\tau) \qquad (\ \dot{f} = df/dt)
\end{aligned}
$$

Problem 18: Using the defining integral (equation 33), demonstrate the five properties listed above. To demonstrate the first property consider the function $f(t) + f(t+\tau)$ and its auto-correlation function properties.

VIII SPECTRAL DENSITY

The spectral density is the basic design tool in the optimization of time-invariant linear systems. The operations of determining the optimum system are reduced to algebraic manipulations by the use of the spectral density. The development of the technique of optimization may well have begun with Wiener's recognition that the spectral density provides a simple link between stationary time series, system transfer functions, and the mean square error of the system. Largely due to the lack of such a tool for the non-stationary case, the design of time-variant systems is both difficult and cumbersome.

A. Definition

The spectral density, $\Phi(\omega)$, of a function f(t) is defined as the average power required at each frequency* in order to describe f(t) in a power** sense. For example, if $f(t) = a \sin(bt + c)$, the spectral density consists of two delta functions of area $a^2/4$ located at $\pm b/2\pi$ on the frequency axis. The major use of spectral density, however, is not the description of known sine waves but of stationary random functions. Such functions possess continuous spectral densities. The spectral density of a random function may be measured directly by scanning the whole frequency region with a narrow band-pass filter. The bandwidth is assumed small

---

* Frequency is defined from minus infinity to plus infinity both for mathematical convenience and because this definition prevents many mistakes in problems involving modulation and detection.

** In this development, "power" is a customary word substitute for "value squared" rather than a measure of work capabilities.

enough so that the power density does not change appreciably within the bandwidth. Inasmuch as the filter scans ± frequencies at the same time, the spectral density at either a positive frequency or its corresponding negative frequency is given by one-half the power output of the filter divided by the filter bandwidth. Figure 11 shows a typical Wiener spectral density and the action of a narrow band scanning filter.

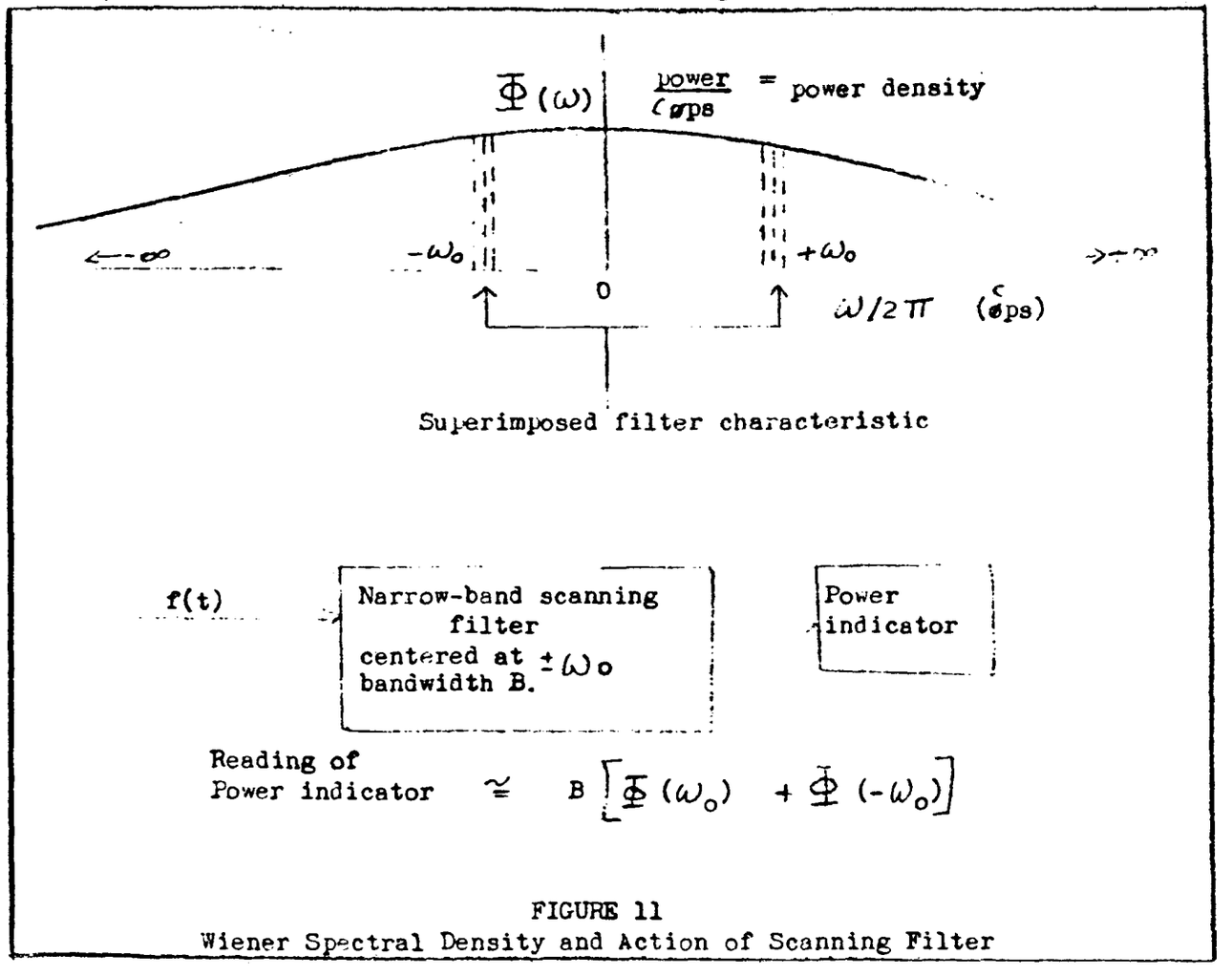


FIGURE 11
Wiener Spectral Density and Action of Scanning Filter

If the density is integrated over all frequencies, then by definition the result is the (average) power of the function f(t).

$$\sigma_f^2 = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{+T} |f(t)|^2 \, dt = \int_{-\infty}^{+\infty} \Phi(\omega) \frac{d\omega}{2\pi} \qquad (35)$$

$$= \frac{1}{2\pi j} \int_{-j\infty}^{+j\infty} \Phi(s) \, ds$$

B. Properties

Perhaps Wiener's greatest contribution was the rigorous demonstration that, for stationary time series, the spectral density and the correlation function are $\mathcal{F}$-transforms of each other. This single contribution changed the spectral density from the level of empirical graphs to the level of analytical mathematics. Properties of Fourier transforms become properties of the spectral density. Referring to the sections on Fourier transforms and correlation functions we can write:

$$\begin{aligned} \Phi_{ff}(s) &= \int_{-\infty}^{+\infty} \varphi_{ff}(\tau)\, e^{-s\tau}\, d\tau \\ \varphi_{ff}(\tau) &= \frac{1}{2\pi j} \int_{-j\infty}^{+j\infty} \Phi_{ff}(s)\, e^{s\tau}\, ds \end{aligned} \qquad (36)$$

Another concept is almost immediately introduced which might otherwise be missed intuitively: the so-called cross-spectral density, a power density present due to relatedness between two time functions.

$$\begin{aligned} \Phi_{fg}(s) &= \int_{-\infty}^{\infty} \varphi_{fg}(\tau)\, e^{-s\tau}\, d\tau \\ \Phi_{gf}(s) &= \int_{-\infty}^{\infty} \varphi_{gf}(\tau)\, e^{-s\tau}\, d\tau \end{aligned} \qquad (37)$$

Based on the properties of the correlation function, we can now write a few properties of the spectral density:

$$1.\quad \Phi_{ff}(s) = \Phi_{ff}(-s)$$

$$2.\quad \sigma_f^2 = \varphi_{ff}(0) = \frac{1}{2\pi j}\int_{-j\infty}^{+j\infty} \Phi_{ff}(s)\, ds$$

$$3.\quad \Phi_{fg}(s) = \Phi_{gf}(-s)$$

$$4.\quad \frac{d}{d\tau}\varphi_{fg}(\tau) = \frac{1}{2\pi j}\int_{-j\infty}^{+j\infty} s\,\Phi_{fg}(s)\, e^{s\tau}\, ds \qquad (38)$$

$$5.\quad \frac{d^n}{d\tau^n}\varphi_{fg}(0) = \frac{1}{2\pi j}\int_{-j\infty}^{+j\infty} s^n\,\Phi_{fg}(s)\, ds$$

Problem 19: Demonstrate the above properties.

Using the notation $F_T(s)$ for the finite-time $\mathcal{L}$ transform of $f(t)$ and $G_T(s)$ for the equivalent in $g(t)$

$$\Phi_{gf}(s) = \mathcal{L}\left[\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{+T} f(t)\, g(t+\tau)\, dt\right]^*$$

$$= \lim_{T\to\infty}\frac{1}{2T}\, F(-s)\, G(s) \qquad (39)$$

$$\Phi_{ff}(s) = \lim_{T\to\infty}\frac{1}{2T}\left|F_T(s)\right|^2$$

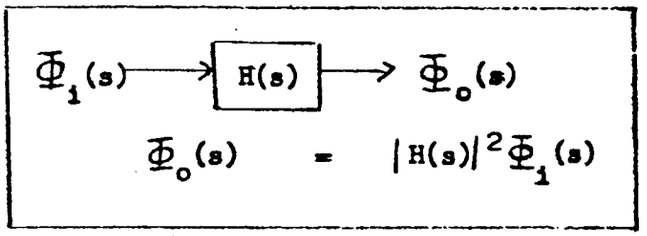
These equations must be used with extreme caution. The cross-spectral density

---

*By convention, the order of subscripts gf defines the F(-s)G(+s) transform combination. The order, fg, would refer to F(+s)G(-s).

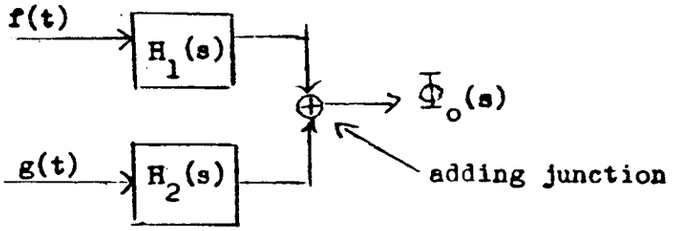
equation is particularly treacherous. The finite-time transforms given here must all be taken with respect to the same time reference. With proper caution, however, it is possible to derive the following properties of spectra passed through systems.

1



$$\Phi_o(s) = |H(s)|^2 \Phi_i(s) \qquad (40)$$

2

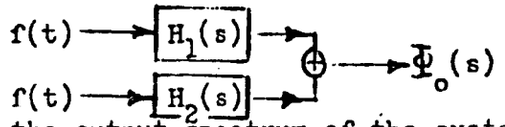


$$\Phi_o(s) = |H_1(s)|^2 \Phi_{ff}(s) + |H_2(s)|^2 \Phi_{gg}(s) + H_1(s)\, H_2(-s) \Phi_{fg}(s) + H_1(-s) H_2(s) \Phi_{gf}(s) \qquad (41)$$

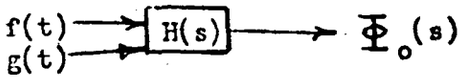
Generally speaking, the amount of cross-correlation is small between disturbances or between disturbances and signals. If the various time functions are independent of each other, the cross-spectral densities are identically zero. The converse is not necessarily true.

Problem 20: What is the output spectral density of a realizable network whose function is $\exp(-\alpha\tau)$ to a stationary time series of autocorrelation function $\exp(-\beta|\tau|) \cdot \cos \gamma\tau$?

Problem 21: Find the output spectrum of the system given below by first combining the transfer functions. Compare with equation (41).



Find the output spectrum of the system given below. Compare with both equation 41 and equation 40 (addition at input).



Applying the Wiener-Paley criterion

$$\frac{1}{2\pi j}\int_{-j\infty}^{j\infty}\left|\frac{\log|\Phi(s)|}{1-s^2}\right| ds < \infty \qquad (42)$$

to a given spectral density determines its factorability, i.e., whether it can be written as

$$\Phi(s) = \Psi(s)\,\Psi(-s) \qquad (43)$$

Wiener has shown that when the criterion does not hold, the stationary time series under consideration is completely predictable. Two such spectral densities are $\Phi(s) = e^{s^2}$, and $e^{-|s|}$, with correlation functions of the form $Ae^{(\tau^2)}$ and $A/(1+\tau^2)$ respectively.

Normal spectral densities, however, may usually be represented as the quotient of two polynomials in s^2.

$$\Phi_{ff}(s) = \frac{a_o - a_1 s^2 + a_2 s^4 - \ldots + \ldots s^{2m}}{b_o - b_1 s^2 + b_2 s^4 - \ldots + \ldots s^{2n}} \qquad n > m$$

or

$$\Phi_{ff}(\omega) = \frac{a_o + a_1\omega^2 + \ldots\ldots + \ldots\ldots \omega^{2m}}{b_0 + b_1\omega^2 + \ldots\ldots + \ldots\ldots \omega^{2n}} \qquad (44)$$

where $s = j\omega$ and where all coefficients are real and positive.

Such spectral density forms are factorable into

$$\Phi_{ff}(s) = \left(\frac{\alpha_0 + \alpha_1 s + \ldots\ldots + \alpha_m s^m}{\beta_0 + \beta_1 s + \ldots\ldots + \beta_n s^n}\right)\left(\frac{\alpha_0 - \alpha_1 s + \ldots\ldots + \alpha_m s^m}{\beta_0 - \beta_1 s + \ldots\ldots + \beta_n s^n}\right) \quad (45)$$
$$= \Psi(s)\,\Psi(-s)$$

where $\Psi(s)$ contains poles and zeros in the left half-plane only, and $\Psi(-s)$ contains their mirror images in the right hand plane.

The correlation functions corresponding to these spectral densities are sums of exponential time functions with complex exponents. The familiar relationships of Fourier transforms to each other hold as well between a function's spectral density and its correlation function. The correlation behavior near $\tau = 0$ describes spectral behavior as $s \to \omega$. Transforms occur in "pairs," i.e., except for the constant 2π, transforming from the τ domain to the ω domain is the same as transforming from the ω domain to the $(-\tau)$ domain. For example:

1. $$\mathcal{F}\left[e^{-a|t|}\right] = \frac{2a}{a^2 + \omega^2}$$

$$\mathcal{F}^{-1}\left[e^{-a|\omega|}\right] = \frac{2a}{2\pi(a^2 + \tau^2)}$$

2. $$\mathcal{F}\left[\delta(t)\right] = 1$$

$$\mathcal{F}^{-1}\left[\delta(\omega)\right] = 1/2\pi$$

C. Comparison with the Fourier "power spectrum"

The definition of spectral density sounds much like the definition of "power" present in a Fourier transform. The question is often asked, if f(t) has the transform F(s) what is the relationship of $|F(s)|^2$ to the spectral density? The answer to the question as stated is: none. From equation 39:

$$\begin{aligned} \Phi_{ff}(s) &= \lim_{T\to\infty} \left[\frac{1}{2T} \left| F_T(s) \right|^2 \right] \\ &= \left| F(s) \right|^2 \lim_{T\to\infty} \left[\frac{1}{2T} \right] \\ &= 0 \end{aligned} \tag{46}$$

Functions which possess Fourier transforms evidently possess no spectral density. Such functions also have zero mean square value. To be transformable f(t) obeys

$$\int_{-\infty}^{+\infty} \left| f(t) \right|^2 dt \quad \begin{matrix} < \infty \\ = A \end{matrix} \tag{23}$$

thus

$$\sigma_f^2 = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{+T} \left| f(t) \right|^2 dt = \frac{A}{\infty} = 0$$

As will be seen in the next section, the Fourier power spectrum can be used to generate a spectral density by introducing randomness. The normal Fourier transform, however, is evidently not the means by which the mean square value of a function is determined.

IX THE FLAT NOISE CONCEPT

Later sections will show that if the spectral densities of all input functions are known, the optimum system may be specified by a few algebraic manipulations. The design engineer's major problem is thus the specification of input spectral densities. The abstraction, "flat noise," is often useful in this connection.

Flat noise is defined as a random time function whose spectral density is a constant.

$$\Phi_{\text{flat}}(\omega) = N^2$$

The correlation function of such a spectral density is $\phi(\tau)$, a delta function at $\tau = 0$. A flat spectral density can be produced by a time-sequence of independent impulses* having zero average value, a mean square area of N^2, and occurring at random time intervals of mean length ΔT_N. For this model of flat noise,

$$\Phi_{\text{flat}} = N^2 = N^2/\Delta T_F$$

Flat noise requires infinite average power. As will be seen, however, the abstraction is often useful.

The first use of flat noise is in the specification of thermal or vacuum tube noise. Such noise has a constant spectral density by actual measurement well past 100,000 megacycles. For all practical purposes, therefore, thermal noise is flat. It is tacitly assumed that the density finally decreases to zero in order that infinite power is not required. For thermal noise, the spectral density of voltage is

---

*Theory of Servomechanisms, pages 298-299

$$\Phi(\omega) = 2RKT$$

where R = resistance in ohms

k = Boltzman's constant

T = absolute temperature in degrees Kelvin

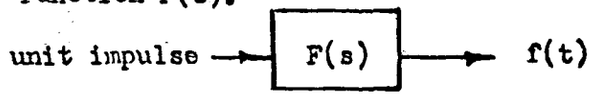
The second use is for mathematical convenience only. We usually know the approximate signal bandwidths. The bandwidth of any circuit we might design certainly will not exceed the signal bandwidths by much more than an order of magnitude. Thus if the noise is flat out to at least 10 signal bandwidths its behavior outside such limits will not be important in specifying the optimum circuit. For example, suppose that the signals have a bandwidth ω_o centered at zero frequency, and that the noise spectral density is given by

$$\Phi_{noise} = \frac{N^2}{1 + \left[\omega/10\omega_o\right]^2}$$

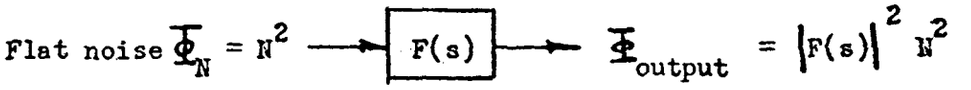
This expression could be used throughout the algebraic manipulations. Generally speaking, the resulting optimum system will not be affected noticably by ignoring the $\left[\omega/10\omega_o\right]^2$ term. For this reason, flat noise is often used as a first approximation in the description of wide-bandwidth disturbances.

The third use of flat noise is often the most helpful. We noticed that Fourier transforms of time functions were unsatisfactory in describing spectral densities or mean square values. The behavior of systems to spectral densities, however, suggests an interesting way of generating spectral densities for signals (and disturbances) whose approximate time characteristics are known. Assume a fair knowledge of f(t) and hence F(s). From transfer function

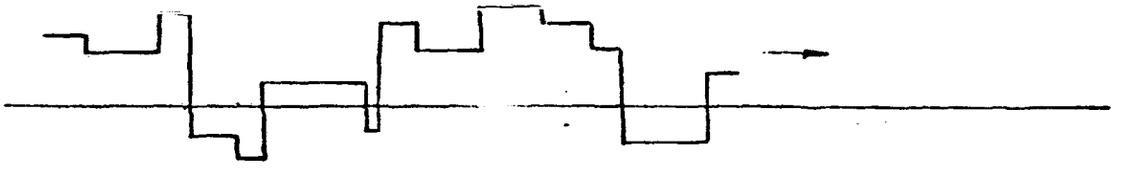
theory we know that f(t) can be produced by inserting a unit impulse into a circuit of transfer function F(s).



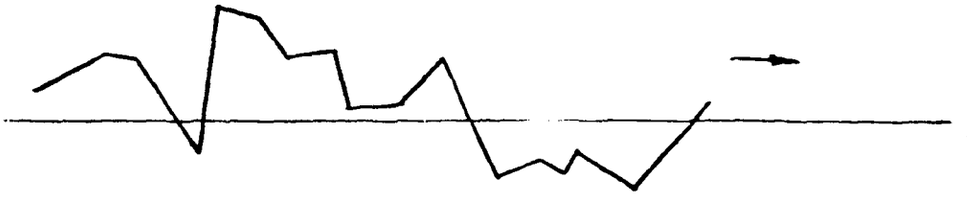
Let us substitute flat noise for the impulse. The output spectral density is $|F(s)|^2 N^2$



The output time function of this system consists of many f(t)'s of random amplitude and time of occurrence. The multiple f(t)'s will overlap, add, and cancel. For example, if f(t) had been assumed a step function, the noise-driven circuit would produce *



or if f(t) = t, the circuit would produce



* Using the random impulse model for flat noise and assuming that the length of the experiment in time is not infinite. Given infinite time, it can be shown that both functions above will drift to either + or - ∞.

The Laplace transforms of a step function and a ramp function are $1/s$ and $1/s^2$, respectively. The equivalent circuits F(s) for the illustrative f(t) are therefore an integrator and a double integrator. The sketches above thus represent integrated and doubly-integrated flat noise. Other functions of time result in appropriate circuits. Occasionally circuits are required with infinite power capabilities. Because of the demonstrable properties of integrated flat noise to drift to plus or minus infinity, both the integrator and double integrator fall into this class. The issue may be side-stepped by stating that there is no such thing as a perfect integrator and that all integrators should be represented by $1/s+a$ where a is some very small number. If in doubt about the mathematical validity of certain operations involving such circuits, deliberately degrade the circuit during the optimization process, removing the degradation at the end of the process if possible.

Several important points arise in this last use of flat noise. The first point deals with the s- and t-symmetry of the final spectral density. Inasmuch as the spectral density is $N^2\ F(s)F(-s)$, the generating time function may have been either f(t) or f(-t). If f(t) were e^{-at}, for example, the spectral density describes not only a random series of e^{-at} but $e^{-a|-t|}$ as well. In addition, flat noise consists of other time functions than a random set of impulses, as may readily ascertained by viewing wide-band noise on an oscilloscope. The preceeding sketches, therefore, describe <u>possible</u> outputs only.

The design engineer completes the generation of spectral densities from approximately known f(t)'s by specifying the expected mean square value of the f(t) and its repetition period. If the mean square value is σ_f^2

and the mean repetition rate ΔT_f, then the spectral density for a random series of f(t)'s is given by

$$\Phi(\omega) = \sigma_F^2 \, \Delta T_f^{-1} \left| F(s) \right|^2$$

Problem 22: A rocket is fired vertically into the atmosphere. Wind is an important disturbance. Assuming that the atmosphere is characterized by layers of random, constant-velocity winds and that the rocket encounters new layers at a roughly constant rate, derive a spectral density for the wind velocity. Assuming that the rocket has a transfer function $1/Ts^2$ from wind velocity to missile position, what is the spectral density of rocket position?

Problem 23: Derive the spectral density of the vertical acceleration produced by bumps in a road. The bump height y is given approximately by ax^2e^{-bx} where x is horizontal distance. Assume constant velocity travel down the bumpy road.

Specification of the spectral density of disturbances should be made at the origin of the disturbance. Specification of the tracking noise of a radar, for example, is most accurately obtained by locating the sources of the noise within the tracking loop, specifying the noise spectral densities at the source, and then using the known transfer characteristics of the tracking loop to specify tracking noise. It is surprising how often a flat noise assumption for the source noise will yield a spectral density which checks empirically at the point of interest in a system. Tracing the source of the noise will occasionally demonstrate certain cross-spectral densities as well.

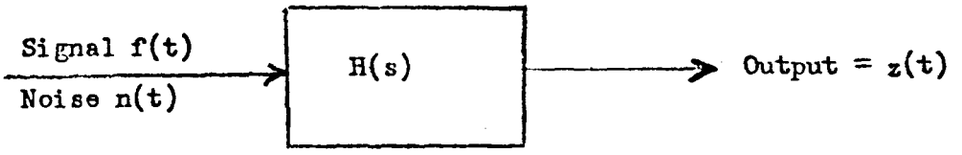
X THE ERROR SPECTRAL DENSITY

Now we begin to hit pay dirt. The desired operation of the system is known. The input spectral densities are known. By writing an expression for the system error (actual performance minus desired performance) converting this expression into its spectral density form $\Phi_{error}(s)$, we may specify the mean square error of the system.

$$\sigma^2_{system} = \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} \Phi_{\varepsilon}(s)\, ds \qquad (47)$$

Given such an integral expression for mean square error, we may minimize it by any technique applicable.

Consider the basic system given in figure 3.



Describing the output by a Laplace transform equation:

$$\mathcal{Z}(s) = H(s)\, F(s) + H(s)\, N(s)$$

Assume that the desired operation is to obtain the best f(t) possible.*

The error of the system is thus

$$\begin{aligned} \mathcal{E}(s) &= \mathcal{Z}(s) - F(s) \\ &= \left[H(s) - 1 \right] F(s) + H(s)\, N(s) \end{aligned} \qquad (48)$$

The squared error is given by

---

* The so-called smoothing operation.

$$\mathcal{E}(s)\,\mathcal{E}(-s) = |H(s) - 1|^2 F(s) F(-s) + |H(s)|^2 N(s) N(-s) + [H(s) - 1] H(-s) F(s) N(-s) + [H(-s) - 1] H(s) F(-s) N(s) \tag{49}$$

The Fourier Transforms F(s) and N(s) do not actually exist for stationary time series. The error-squared expression of equation 49 is used only as a shortcut method of finding the appropriate spectral densities. Equation 49 could be justified by considering F(s) and N(s) as finite-time transforms. The spectral density form of equation 49 may therefore be written by taking the necessary time averages (see equation 39).

$$\overline{\Phi}_{\mathcal{E}}(s) = |H(s) - 1|^2 \overline{\Phi}_{ff}(s) + |H(s)|^2 \overline{\Phi}_{nn}(s) + [H(s) - 1]\, H(-s)\, \overline{\Phi}_{fn}(s) + [H(-s) - 1]\, H(s)\, \overline{\Phi}_{nf}(s) \tag{49a}$$

If the signal and noise are uncorrelated, both $\overline{\Phi}_{nf}$ and $\overline{\Phi}_{fn}$ are zero. It is <u>not</u> legitimate to find cross-spectral densities by factoring the signal and noise spectral densities into functions of +s and -s which are then combined into apparent cross-spectral densities. Cross-spectral densities must be justified independently of the form of the auto-spectral densities.

<u>Problem 24:</u> Assume a system in which the signal f(t) and noise n(t) arrive at the output by two different routes such that $H_1(s)$ acts on f(t) and $H_2(s)$ on n(t). The desired system operation is $H_3(s)$ acting on f(t). Assuming correlation between signal and noise, what is the error spectral density?

<u>Problem 25:</u> What is the error spectral density of a smoothing system $H(s) = \frac{1}{a + s}$ to a signal of spectral density $(b^2 - s^2)^{-1}$ and a flat noise spectral density? Express the answer as the quotient of two polynomials.

XI PHILLIPS' OPTIMIZATION TECHNIQUE *

If a complete system is at our disposal it is worthwhile to design the optimum possible system. Often, however, the system is completely specified by other considerations and only a few parameters are available which can be varied to yield best results. For example, the optimum system may require isolation amplifiers in order to realize the optimum transfer function. Addition of such amplifiers may so increase the cost that in a competitive market the result is economic loss.

In the previous sections we have shown how to specify the error spectral density for a system and have remarked that spectral densities usually occur as quotients of polynomials. Phillips noted this fact and set about evaluating the integral below in terms of the coefficients a and b.

$$\sigma_n^2 = \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} \frac{g_n(s)}{h_n(s)h_n(-s)}\, ds$$

where

$$h_n(s) = a_0 s^n + a_1 s^{n-1} + \ldots\ldots a_n \qquad (51)$$

$$g_n(s) = b_0 s^{2n-2} + b_1 s^{2n-4} + \ldots\ldots b_{n-1}$$

The roots of $h_n(s)$ must all lie in the left half plane. The factorization of the denominator puts the integral into its most useful form. The numerator is not factored because such factorization, while certainly possible, would entail additional work for the user. The subscript n gives the order of complexity of the polynomials. As an illustration let

---

* From "Theory of Servomechanisms" Vol. 25, MIT Radiation Laboratory Series

$$\Phi_\varepsilon = \frac{A}{\alpha^2 - s^2} + \frac{B}{\beta^2 - s^2}$$ Evaluate directly

$$= \frac{-(A+B)s^2 + (A\beta^2 + B\alpha^2)}{[s^2 + (\alpha+\beta)s + \alpha\beta][s^2 - (\alpha+\beta)s + \alpha\beta]}$$

$$\sigma_\varepsilon^2 = \frac{1}{2\pi j}\int_{-j\infty}^{j\infty} \Phi_\varepsilon \, ds = \frac{A}{2\alpha} + \frac{B}{2\beta}$$

Using Phillips' integral expression,

$$h_n(s) = s^2 + (\alpha+\beta)s + \alpha\beta = h_2(s)$$

$$g_n(s) = -(A+B)s^2 + (A\beta^2 + B\alpha^2) = g_2(s)$$

and thus

$n = 2$
$a_0 = 1$
$a_1 = \alpha + \beta$
$a_2 = \alpha\beta$

$b_0 = -(A+B)$
$b_1 = A\beta^2 + B\alpha^2$

Phillips has evaluated such integrals for values of n form 1 to 7. (Theory of Servomechanisms page 369). The first four integrals are

$$\sigma_1^2 = \frac{b_o}{2a_oa_1}$$

$$\sigma_2^2 = \frac{-b_o + \frac{a_ob_1}{a_2}}{2a_oa_1} \qquad (52)$$

$$\sigma_3^2 = \frac{-a_2b_o + a_ob_1 - \frac{a_oa_1b_2}{a_3}}{2a_o\,(a_oa_3 - a_1a_2)}$$

$$\sigma_4^2 = \frac{b_o(-a_1a_4 + a_2a_3) - a_oa_3b_1 + a_oa_1b_2 + \frac{a_ob_3}{a_4}\,(a_oa_3 - a_1a_2)}{2a_o\,(a_oa_3^2 + a_1^2\,a_4 - a_1\,a_2\,a_3)}$$

The expression for σ_5^2 contains 21 terms. σ_6^2 contains 47 terms and σ_7^2, 113.

Problem 26: What is the mean square error given that

$$\Phi_\varepsilon = \frac{10}{1 - s^2} + \frac{3}{16 - s^2} \text{ volts}^2/\text{cps}$$

The optimization problem in Phillips technique is the problem of minimizing the σ_n^2 with whatever variable parameters are available. For example, let

$$\Phi_\varepsilon = \frac{A}{\alpha^2 - s^2} + \frac{B\,\alpha}{\beta^2 - s^2}$$

where α is variable then

$$\sigma^2 = \frac{A}{2\alpha} + \frac{B\alpha}{2\beta}$$

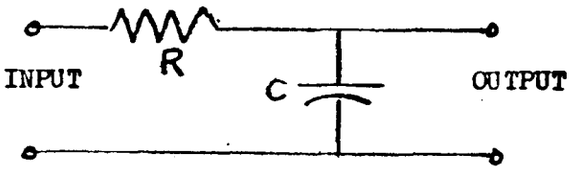
for a minimum

$$\frac{d\sigma^2}{d\alpha} = 0$$

and consequently

$$\alpha = +\sqrt{\frac{A}{B}\beta}$$

Problem 27: What is the best RC time constant in the network below to minimize the mean square error when smoothing a signal $\Phi_{ff} = \frac{2\beta}{\beta^2 - s^2}$ in the presence of (independent)noise $\Phi_{nn} = \frac{2\alpha}{\alpha^2 - s^2}$



This optimization technique followed Wiener's technique by roughly a year, even though, mathematically, this technique is far more limited in scope.

unit power in each. Assume $Y = \frac{1}{1+\tau s}$. Find best τ. $\frac{S}{N} = 1$

$$\phi_\varepsilon = \left| \frac{1}{(1+\tau s)(1-\tau s)} \right| \frac{2\alpha}{(\alpha+s)(\alpha-s)} + \left| \frac{-\tau s}{1+\tau s} \right|^2 \left| \frac{2\beta}{\beta^2 - s^2} \right|$$

= noise term + sig distort.

$$= \frac{2\alpha}{(1+\tau s)(\alpha+s)(\quad)} + \frac{2\beta\tau^2}{(1+\tau s)(\beta+s)}$$

$$= \frac{\alpha + (\tau\alpha + 1)s + \tau s^2}{\quad}$$

$n = 2$ $\quad a_0 = \tau$, $a_1 = 1 + \tau\alpha$, $a_2 = \alpha$

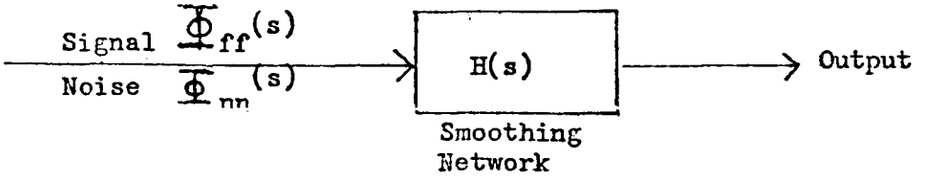
$b_0 = 0$, $b_1 = 2\alpha$

1 st int.

XII WIENER OPTIMUM FILTERS

A. Basic Logic

The Phillips technique minimizes the mean square error of a system in which a limited number of parameters are adjustable. The Wiener optimum is achieved via the calculus of variations route. However, the derivation of the optimum, as Wiener presents it in his "Stationary Time Series," is so mathematical in language as to obscure the simplicity of logic on which it is based. Indeed, the logic has already been presented. Except for a trick necessary to make the system realizable, and by considering an integral without absolute value brackets, we optimized the simple smoothing system in problem 1 a. Combining our knowledge of error spectral densities with that of the calculus of variations we may "almost" find Wiener's answer in three lines.



(Signal and Noise Independent)

FIGURE 12
The Basic Wiener Smoothing Problem

$$\sigma_\varepsilon^2 = \frac{1}{2\pi j}\int_{-j\infty}^{j\infty} \Phi_{\varepsilon\varepsilon}\, ds = \frac{1}{2\pi j}\int_{-j\infty}^{j\infty} \left[\Phi_{NN}\left|H(s)\right|^2 + \Phi_{ff}\left|1-H(s)\right|^2\right] ds$$

$$\overset{\text{"almost"}}{=>} \frac{1}{2\pi j}\int_{-j\infty}^{j\infty} \left[\Phi_{NN}\, H^2 + \Phi_{ff}\,(1-H)^2\right] ds \qquad (53)$$

which is minimized* by the calculus of variations technique, yielding:

$$H(s) = \frac{\Phi_{ff}(s)}{\Phi_{ff}(s) + \Phi_{NN}(s)} \qquad (54)$$

Problem 28: Show that if the desired operation has been $H_1(s)$ instead of smoothing, the above H(s) would have been

$$H(s) = \frac{H_1(s)\,\Phi_{ff}(s)}{\Phi_{ff}(s) + \Phi_{NN}(s)} \qquad (55)$$

and that

$$\sigma^2_{opt} = \frac{1}{2\pi j}\int_{-j\infty}^{j\infty} \frac{|H_1(s)|^2 \Phi_{ff}\,\Phi_{NN}}{\Phi_{ff} + \Phi_{NN}}\,ds \qquad (56)$$

The H(s) which we have just determined is interesting but cannot be realized. As was shown in the section on spectral densities:

$$\Phi_{ff}(s) = \Phi_{ff}(-s)$$
$$\Phi_{NN}(s) = \Phi_{NN}(-s)$$

and consequently H(s) = H(-s). The transfer function thus has terms such as

$$\frac{2\alpha}{\alpha^2 - s^2} = \frac{1}{\alpha + s} + \frac{1}{\alpha - s}$$

the second of which is either unstable $(e^{\alpha t})$ or unrealizable $e^{-\alpha(-t)}$ as was discussed in the section on Fourier transforms.

Inasmuch as this obstacle takes only a little trickery to overcome, we have solved for the famous Wiener optimum filter as far as basic concepts are concerned.

B. Trickery

The purpose of the following trickery is to make the variation of σ_ϵ^2 zero with an H(s) which is realizable. The trickery enables us to discard

*Problem 1a.

part of the integral resulting from the calculus of variations differentiation. Setting the remainder of the integral equal to zero yields the optimum realizable filter.

The essence of the maneuver is contained in the values of two integrals.*

$$I_1 = \frac{1}{2\pi j}\int_{-J\infty}^{J\infty} \frac{ds}{(\alpha+s)(\beta+s)} = \frac{1}{2\pi j}\int_{-J\infty}^{J\infty} \frac{ds}{(\alpha-s)(\beta-s)} = 0$$

$$I_2 = \frac{1}{2\pi j}\int_{-J\infty}^{J\infty} \frac{ds}{(\alpha-s)(\beta+s)} = \frac{1}{2\pi j}\int_{-J\infty}^{J\infty} \frac{ds}{(\alpha+s)(\beta-s)} = \frac{1}{\alpha+\beta} \qquad (57)$$

The constants α and β are complex numbers whose real part is positive.

Let us define two useful forms by a plus and minus subscript such that

$$\mathcal{Z}_+(s) = \sum_i \frac{Ai}{\alpha_i + s}$$
$$\mathcal{Z}_-(s) = \sum_i \frac{Ai}{\alpha i - s} = \mathcal{Z}_+(-s) \qquad (58)$$

The $\mathcal{Z}_+(s)$ is defined as having poles in the left-half plane only. $\mathcal{Z}_-(s)$ has its poles in the right-half plane only. Then if $Y_+(s)$ and $Y_-(s)$ is another such pair of functions, the integrals I_1 and I_2 demonstrate that by pairing terms in $\mathcal{Z}$ and Y,

---

* These values may be obtained most easily by contour integration. If such a technique is unfamiliar to the reader, the values may be accepted on faith. It is probably unprofitable to study the technique just for this development.

$$I_1' = \frac{1}{2\pi j}\int_{-j\infty}^{j\infty} Z_+ Y_+ \, ds = \frac{1}{2\pi j}\int_{-j\infty}^{j\infty} Z_- Y_- \, ds = 0$$

$$I_2' = \frac{1}{2\pi j}\int_{-j\infty}^{j\infty} Z_- Y_+ \, ds = \frac{1}{2\pi j}\int_{-j\infty}^{j\infty} Z_+ Y_- \, ds = \text{Constant} \qquad (59)$$

Equation 59 shows that integrals whose integrands have poles in one half-plane only have value zero. Those with "mixed" poles have a non-zero value.

We now proceed to tamper with the minimization carried out earlier.* Noting that

$$\begin{aligned} |H(s)|^2 &= H_+ H_- \\ |1 - H(s)|^2 &= (1 - H_+)(1 - H_-) \end{aligned} \qquad (60)$$

we rewrite

$$\sigma^2 = \frac{1}{2\pi j}\int_{-j\infty}^{j\infty} \left[\Phi_{NN}|H(s)|^2 + \Phi_{ff}|1-H(s)|^2\right] ds \qquad (61)$$

$$= \frac{1}{2\pi j}\int_{-j\infty}^{j\infty} \left[\Phi_{NN} H_+H_- + \Phi_{ff}(1 - H_+)(1 - H_+)\right] ds$$

To minimize this integral we replace H_+ by $H_+ + \varepsilon\eta_+$. The arbitrary function η_+ is subject to the same restriction as H_+, namely that it be realizable -- all its poles in the left-half plane. Performing the substitution and setting $\partial\sigma^2/\partial\varepsilon = 0$ at $\varepsilon = 0$ yields

*Problem 1 a

$$\left.\frac{\partial \sigma^2}{\partial \varepsilon}\right|_{\varepsilon=0} = \frac{1}{2\pi j}\int_{-j\infty}^{j\infty} \eta_+ \left[H_- (\Phi_{ff} + \Phi_{NN}) - \Phi_{ff}\right] ds$$

$$+ \frac{1}{2\pi j}\int_{-j\infty}^{j\infty} \eta_- \left[H_+ (\Phi_{ff} + \Phi_{NN}) - \Phi_{ff}\right] ds \qquad (62$$

We can obtain a zero value, of course, by

$$H_+ = H_- = \frac{\Phi_{ff}}{\Phi_{ff} + \Phi_{NN}} \qquad (63)$$

but this solution is unsatisfactory. The trick consists of demonstrating that part of the right hand side of the integral equation is zero, regardless of η or the spectral densities. Let us factor $\left[\Phi_{NN} + \Phi_{ff}\right]$

$$\Phi_{NN} + \Phi_{ff} = \Psi_+ \Psi_- \qquad (64)$$

The section on Fourier transforms gives us justification for this action.

Then

$$\left.\frac{\partial \sigma^2}{\partial \varepsilon}\right|_{\varepsilon=0} = \frac{1}{2\pi j}\int_{-J\infty}^{J\infty} \eta_+ \Psi_+ \left[H_- \Psi_- - \frac{\Phi_{ff}}{\Psi_+}\right] ds \qquad (65)$$

$$\frac{1}{2\pi j}\int_{-J\infty}^{J\infty} \eta_- \Psi_- \left[H_+ \Psi_+ - \frac{\Phi_{ff}}{\Psi_-}\right] ds$$

The first parts of these integrals, involving the integrals $\int \eta_+ \Psi_+ H_- \Psi ds$

and $\int \eta_- \Psi_- H_+ \Psi_+ \, ds$, can not be zero because they contain integrands with mixed poles (equation 59). The second parts of the integrals, however, contain both zero and non-zero parts. To show this, split $\frac{\Phi}{\Psi_+}$ and $\frac{\Phi}{\Psi_-}$ into sums of partial fractions

$$\frac{\Phi_{ff}}{\Psi_-} = \sum_i \frac{A_i}{\alpha_i + s} + \sum_j \frac{B_j}{\beta_j - s} = \left(\frac{\Phi_{ff}}{\Psi_-}\right)_+ + \left(\frac{\Phi_{ff}}{\Psi_-}\right)_-$$

$$\frac{\Phi_{ff}}{\Psi_+} = \sum_i \frac{A_i}{\alpha_i - s} + \sum_j \frac{B_j}{\beta_j + s} = \left(\frac{\Phi_{ff}}{\Psi_+}\right)_- + \left(\frac{\Phi_{ff}}{\Psi_+}\right)_+$$

The zero parts of the integrals of equation 65 are thus

$$\begin{aligned} 0 &= \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} \eta_+ \Psi_+ \left(\frac{\Phi_{ff}}{\Psi_+}\right)_+ \, ds \\ 0 &= \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} \eta_- \Psi_- \left(\frac{\Phi_{ff}}{\Psi_-}\right)_- \, ds \end{aligned} \tag{67}$$

because the integrands contain poles in one half-plane only. Rewriting equation 65 without its zero parts we obtain

$$\left.\frac{\partial \sigma^2}{\partial \varepsilon}\right|_{\varepsilon=0} = \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} \eta_+ \Psi_+ \left[H_- \Psi_- - \left(\frac{\Phi_{ff}}{\Psi_+}\right)_-\right] ds + \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} \eta_- \Psi_- \left[H_+ \Psi_- - \left(\frac{\Phi_{ff}}{\Psi_-}\right)_+\right] ds \tag{65a}$$

The conditions for zero variation, independent of η, are therefore

$$H_+ \Psi_+ = \left(\frac{\Phi_{ff}}{\Psi_-} \right)_+ ; \; H_- \Psi_- = \left(\frac{\Phi_{ff}}{\Psi_+} \right)_- \tag{68}$$

These equations are exactly alike except for a $\pm s$ interchange. By convention, $(\Phi_{ff}/\Psi_-)_+$ is called the physically-realizable part of Φ_{ff}/Ψ_- and is usually written as

$$\left(\frac{\Phi_{ff}}{\Psi_-} \right)_+ = \left(\frac{\Phi_{ff}}{\Psi(-s)} \right)_{\substack{\text{physically} \\ \text{realizable}}} = \sum \frac{A_i}{\alpha_i + s}$$

The optimum Wiener filter is therefore given by

$$H(s) = \frac{1}{\Psi(s)} \left(\frac{\Phi_{ff}}{\Psi(-s)} \right)_{\substack{\text{physically} \\ \text{realizable}}} \tag{69}$$

Problem 29: Prove that if the desired operation had been $H_1(s)$ instead of smoothing, the optimum system would be

$$H(s) = \frac{1}{\Psi(s)} \left(\frac{H_1(s)\, \Phi_{ff}}{\Psi(-s)} \right)_{\substack{\text{physically} \\ \text{realizable}}} \tag{70}$$

The derivation has applied to a signal and noise which are uncorrelated. In this case of correlated signal and noise the answer is

$$H(s) = \frac{1}{\Psi(s)} \left(\frac{H_1(s)\left[\Phi_{ff} + \Phi_{fn} \right]}{\Psi(-s)} \right)_{\substack{\text{physically} \\ \text{realizable}}} \tag{71}$$

where

$$\Psi(s)\Psi(-s) = \Phi_{ff} + \Phi_{nn} + \Phi_{fn} + \Phi_{nf} \tag{72}$$

$H_1(s)$ = desired minimum phase operation

Problem 30: What are the optimum filters in the cases listed below? Signal and noise are uncorrelated.

Φ_{ff}	Φ_{nn}	$H_1(s)$
$\frac{2\alpha}{\alpha^2 - s^2}$	N^2	1
$\frac{2\alpha}{\alpha^2 - s^2}$	N^2	$1/s$
$\frac{2\alpha}{\alpha^2 - s^2}$	$\frac{2\beta}{\beta^2 - s^2}$	1
$\frac{2\alpha}{\alpha^2 - s^2}$	$\frac{2\beta}{\beta^2 - s^2}$	$\frac{1}{s+\gamma}$

C. How to Specify a Wiener Optimum System

The foregoing proof has demonstrated that the optimum Wiener system is given by

$$H(s) = \frac{1}{\Psi(s)} \left(\frac{H_1(s)\,(\Phi_{ff} + \Phi_{fn})}{\Psi(-s)} \right)_{\substack{\text{Physically} \\ \text{Realizable}}} \tag{71}$$

where $\Psi(s)\,\Psi(-s) = \Phi_{ff} + \Phi_{nn} + \Phi_{fn} + \Phi_{nf}$

$H_1(s)$ = Desired System Operation (72)

and where, if the brackets are expressible

$$\left(\quad \right)_{P.R.} = \left(\frac{a_o + a_1 s + \cdots\cdots s^{K+L-1}}{(\alpha_o + s)(\alpha_1 + s)\cdots(\alpha_K + s)(\beta_o - s)(\beta_1 - s)\cdots(\beta_L - s)} \right)_{P.R.} \tag{73}$$

then

$$\left(\quad \right)_{\substack{\text{Physically} \\ \text{Realizable}}} = \sum_{i}^{K} \frac{A_i}{\alpha_i + s} \tag{74}$$

where the A_i are determined by any one of a number of techniques. (Gardner & Barnes page 164).

D. A Few Cautions in the Use of the Wiener Optimum

1. Perfect prediction (past and future)

The perfect predictor is given by

$$e^{sT} = 1 + sT + \frac{s^2T^2}{2!} + \cdots\cdots \frac{s^nT^n}{n!}, \cdots$$

$$\approx \left(\frac{2n + sT}{2n - sT}\right)^n \tag{75}$$

If T is a positive number (future prediction), the use of $e^{Ts} = H_1(s)$ in the Wiener optimum H(s) equation will contribute little other than effecting the coefficients in the sum

$$\left(\frac{H_1(s)\ (\Phi_{ff} + \Phi_{fn})}{\Psi(-s)}\right)_{P.R.} = \sum_i \frac{A_i}{\alpha_1 + s}$$

because no poles of e^{Ts} appear in the left half-plane. If T is a negative number, however,

$$e^{-s|T|} = \frac{1}{1 + s|T| + \frac{s^2|T|^2}{2!} + \cdots} \tag{76}$$

$$\approx \left(\frac{2n - s|T|}{2n + s|T|}\right)^n$$

and it becomes necessary to include more terms in the operation "physically realizable." Depending upon the approximation used for

$e^{-s|T|}$ we will get different contributions. For example, if the approximation n = 2 is used, from equation 76,

$$\left(\frac{e^{-s|T|}\Phi_{ff} + \Phi_{fn}}{\Psi(-s)}\right)_{P.R.} = \sum_i \frac{A_i}{\alpha_i + s} + \frac{B_1}{(4+s|T|)^2} + \frac{B_2}{4+s|T|} \quad (77)$$

The approximation might also have been used

$$H_1(s) = \frac{1}{1 + |T|\, s + \frac{|T|^2}{2} s^2}$$

resulting in still different poles. The best approximation probably depends upon the signal and noise functions considered.

2. Misleading Operations

The operation, "physically realizable," has been defined only for quotients of polynomials whose numerator is of lesser order than the denominator. Operations $H_1(s)$ which result in

$$\frac{H_1(s)\,(\Phi_{ff} + \Phi_{fn})}{\Psi(-s)}$$

not having a higher order denominator than numerator generally produce misleading results. One simple check before starting the problem will prevent such occurrences. Check that if no noise were present, we would not be performing an operation yielding infinite output signal values. As an example, asking the differentiation of

$$\Phi_{ff} = \frac{1}{1 - s^2}$$

would result in an output spectrum

$$\Phi_{out} = \frac{-s^2}{1 - s^2}$$

whose mean square value is infinite. Stated mathematically, the function f(t) represented by such a Φ_{ff} is almost never differentiable.

Problem 31: a. Find the operator which appears when

$$H_1(s) = s \qquad \Phi_{nn} = N^2$$

$$\Phi_{ff} = \frac{1}{1 - s^2}$$

b. Let $N \rightarrow 0$. Are we differentiating?

E. THE WIENER SOLUTION IN THE TIME DOMAIN

The following time-domain solution is unnecessary in discussing time-invariant systems. It is presented here only to demonstrate the approach to time-variant systems. The time-variant solution is much like the spectral density solution in having a simple logic but requiring some mathematical trickery. We will omit the trickery, referring the reader to Wiener, pages 84-86, for the stationary case. No equivalent trickery exists for the time-variant case; a general solution apparently is impossible.

Let us solve the smoothing problem where the desired result is f(t) and the input is f(t) + n(t), correlated. The output of the linear filter $h(\tau, t)$ is

$$e_{out}(t) = \int_0^\infty h(\tau, t) \left[f(t - \tau) + n(t - \tau) \right] d\tau \qquad (78)$$

The weighting function $h(\tau, t)$ is a slightly more generalized function than usually seen in such integrals. The dependence on time of the weighting function is most easily pictured by imagining the weighting function as changing with time as it is dragged along the time axis.

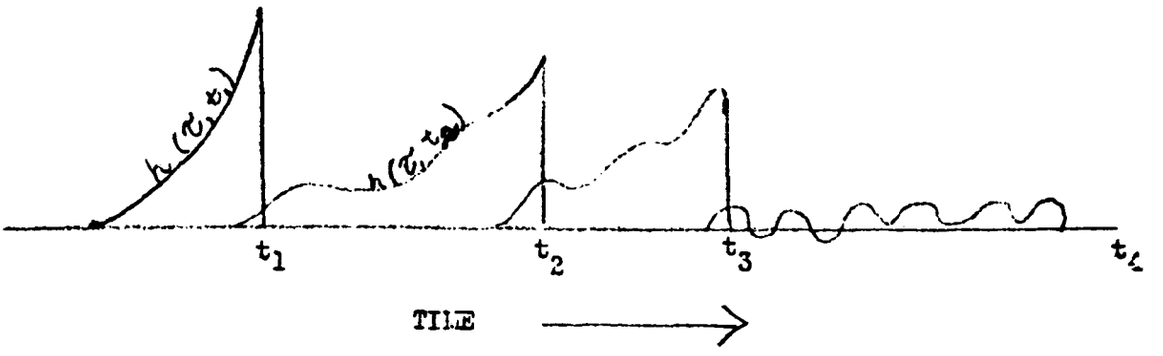


FIGURE 13

A Time-variant Weighting Function

The error of the system at time t is given by

$$\mathcal{E}(t) = e_{out}(t) - f(t)$$

$$= \int_0^\infty k(\tau, t)\left[f(t-\tau) + n(t-\tau)\right] d\tau \quad - f(t) \qquad (79)$$

we now must make an important decision. We must define what is meant by mean in the term mean square error. The definition which has been used to this point (and which we will arbitrarily continue to use) is that the mean is taken with respect to an ensemble of experiments. The mean square error might also have been defined as a time average during one experiment. The latter definition was used in the curve fitting problem. If the systems are stationary the ergodic hypothesis states that the definitions are equivalent. In the system is time-variant the two definitions are not equivalent.

To illustrate the difference between the definitions, let us find the mean square value of the non-stationary time series $e^{-\alpha|t|}n(t)$ where $n(t)$ is a stationary random function of mean zero and mean square value σ_n^2.

Definition One (Ensemble of Experiments)

$$\sigma^2(t) = \overline{\left(e^{-\alpha|t|}\right)^2 n^2(t)}$$

$$= e^{-2\alpha|t|}\sigma_N^2 \qquad (80)$$

Definition Two (Time Average)

$$\sigma^2(t) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} e^{-2\alpha|t|} n^2(t)dt$$

$$= 0$$

Using the ensemble definition leads to the assumption that the mean square error of a system is minimized if it is ensemble-minimized at every instant.

Coming back to the derivation, the mean square (ensemble) average of $\varepsilon(t)$ is

$$\overline{\varepsilon^2(t)} = \overline{\left\{\int_0^\infty h(\tau_1,t)\left[f(t-\tau_1)+n(t-\tau_1)\right]d\tau_1 - f(t)\right\}} \qquad (81)$$
$$\times \overline{\left\{\int_0^\infty h(\tau_2,t)\left[f(t-\tau_2)+n(t-\tau_2)\right]d\tau_2 - f(t)\right\}}$$

which may be written as

$$\overline{\varepsilon^2(t)} = \overline{f^2(t)} - \overline{2f(t)\int_0^\infty h(\tau,t)\left[f(t-\tau)+n(t-\tau)\right]d\tau} \qquad (82)$$
$$+ \overline{\int_0^\infty h(\tau_1,t)\left\{\int_0^\infty h(\tau_2,t)\left[f(t-\tau_1)+n(t-\tau_1)\right]\left[f(t-\tau_2)+n(t-\tau_2)\right]d\tau_2\right\}d\tau_1}$$

Defining some ensemble correlation functions

$$\varphi_{ff}(t,\tau) = \overline{f(t)\,f(t+\tau)}$$
$$\qquad (83)$$
$$\varphi_{fn}(t,\tau) = \overline{f(t+\tau)\,n(t)}$$

we rewrite the mean square error as

$$\overline{\varepsilon^2(t)} = \overline{f^2(t)} - 2\int_0^\infty h(\tau,t)\left[\varphi_{ff}(t,-\tau) + \varphi_{nf}(t,-\tau)\right] d\tau + \int_0^\infty h(\tau_1,t)d\tau_1 \int_0^\infty h(\tau_2,t)\left[\varphi_{ff}(t,\tau_1-\tau_2) + \varphi_{fn}(t,\tau_1-\tau_2) + \varphi_{nf}(t,\tau_1-\tau_2) + \varphi_{nn}(t,\tau_1-\tau_2)\right] d\tau_2 \qquad (84)$$

The mean square error is now minimized by the calculus of variations approach. Let $h(\tau,t)$ be replaced by $h(\tau,t) + \varepsilon\eta(\tau,t)$, differentiate with respect to ε at $\varepsilon = 0$, and make the result independent of $\eta(\tau,t)$. The result is the equation

$$\varphi_{ff}(t,\tau) + \varphi_{fn}(t,\tau) = \int_0^\infty \underbrace{\left[\varphi_{ff} + \varphi_{fn} + \varphi_{nf} + \varphi_{nn}\right]}_{\text{Argument} = t,\ \tau+\tau_1} h(t,\tau_1)d\tau_1 \qquad (85)$$

which holds (only) for $\tau > 0$. This equation must now be solved for $h(t,\tau)$. No solution has been proposed for the time-variant case. For the stationary case where the operation $H_1(s)$ is desired, it is possible to show that

$$H(s) = \frac{1}{2\pi j \Psi(s)} \int_0^\infty e^{-st}dt \int_{-j\infty}^{j\infty} \frac{H_1(s)(\Phi_{ff} + \Phi_{fN})}{\Psi(-s)} ds \qquad (86)$$

which is the same as

$$H(s) = \frac{1}{\Psi(s)} \left(\frac{H_1(s)(\Phi_{ff} + \Phi_{fN})}{\Psi(-s)} \right)_{\text{Physically Realizable}} \qquad (87)$$

derived earlier. From our ensemble definition of mean square error, $H_1(s)$ may be a time-variant operator if desired. A time-variant operator, $H(s,t)$, will appear much like the usual constant-coefficient $H(s)$ but will have t as a parameter.

Problem 32: Prove equation 85 in the manner outlined in the text.

XIII A PHYSICAL INTERPRETATION OF THE WIENER OPTIMUM

Bode and Shannon* have given an instructive picture of the significance of the operations used in obtaining Wiener's optimum system. Their derivation is basically the same as the one presented here. Their approach and language are both slightly different. Unfortunately, they neglected to write the H(s) equation in conclusion.

The engineer recognizes the necessity of making his system physically realizable. He also agrees that

$$H(s) = \frac{\Phi_{ff}}{\Phi_{ff} + \Phi_{nn}} \qquad \text{(Example only)}$$

is a fine solution but unrealizable. But why can't we write the solution as follows?

$$H(s) = \left(\frac{\Phi_{ff}}{\Phi_{ff} + \Phi_{nn}} \right) \qquad \text{(Example Only)}$$

Physically Realizable

The reason, mathematically, is that the optimization derivation does not work this way. The physical reason is this: the total input function (signal plus noise) possesses correlation, i.e., some of the data in the future is predictable and hence we should use a weighting function which understands this fact. The future information will not be totally surprising. Thus, even though our weighting function can only use available data, it can do a better smoothing job by doing some predicting, as well.

---

* April 1950 IRE page 417

On the other hand, suppose the total input spectral density is flat (correlation only at $\tau = 0$). The future is a complete surprise. Under these conditions let us derive a system with no holds barred. The result will be unrealizable as always. But the unrealizable part, the part using future data, is working with completely unpredictable information (correlation of future input data with past input data is zero). The best guess at future behavior of the input is that it will be zero. Because the future is unavailable, we will lose little on the average by assigning it a zero value. The "physically realizable" operation thus costs us the least if performed on the transfer function of a network whose total input spectral density is flat. Let us apply this reasoning to the Wiener problem.

1. Convert the given input to flat noise. The operator is $1/\Psi(s)$.
2. Optimize the system from this point with the desired operation being $\Psi(s)$.

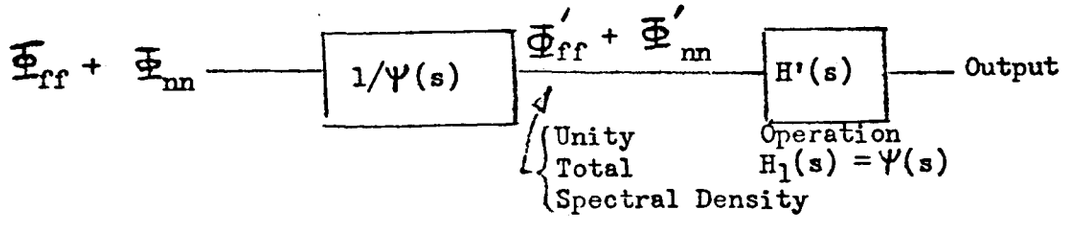


FIGURE 14

Physical System for Finding Wiener Optimum System

The optimum H'(s), no holds barred, is

$$H'(s) = \frac{\Psi(s)\ \Phi'_{ff}}{\Phi'_{ff} + \Phi'_{nn}} \tag{88}$$

as shown earlier. But this may be written as

$$H'(s) = \frac{\Psi(s)\Phi_{ff}/\Psi(s)\Psi(-s)}{(\Phi_{ff}+\Phi_{nn})/\Psi(s)\Psi(-s)} = \frac{\Phi_{ff}}{\Psi(-s)} \qquad (89)$$

3. Taking the physically realizable part and combining into one transfer function

$$H(s) = \frac{1}{\Psi(s)}\left(\frac{\Phi_{ff}(s)}{\Psi(-s)}\right)_{\substack{\text{Physically}\\ \text{Realizable}}} \qquad (90)$$

Problem 33: Derive the optimum H(s) to perform an operation $H_1(s)$ using the same reasoning.

Several consequences of the Wiener solution are evident. First, and probably most important, the best way of performing an operation $H_1(s)$ is not to first smooth the input and then operate with $H_1(s)$. It is not true, for example, that the best derivative of a function is obtained by finding the best smoothed value and then differentiating it. The second consequence is more mathematical than practical. The required correlations are not the signal and noise correlations but the total correlation function of the input and the correlation of the signal on the total input.

$$\varphi_{\substack{\text{Input,}\\ \text{Input}}} = \varphi_{ff} + \varphi_{nn} + \varphi_{fn} + \varphi_{nf}$$

$$\varphi_{\substack{\text{Signal,}\\ \text{Input}}} = \varphi_{ff} + \varphi_{fn}$$

Problem 34: Given $\Phi_{ff} = \frac{1}{\alpha^2 - s^2}$ and $\Phi_{nn} = \frac{1}{\beta^2 - s^2}$

show that the true optimum system differs from $\left(\frac{\Phi_{ff}}{\Phi_{ff}+\Phi_{nn}}\right)$ P.R.

in its numerator prediction time constant.

XIV EXTENSIONS OF WIENER'S SOLUTION

The basic theory of the optimization technique is now complete. The remainder of this text consists of various applications of the Wiener solution to special problems. It is the purpose of this section to describe the various problems. The reader may then select the topics of particular interest in his own problem. A servomechanisms engineer, for example, would probably select the sections on quasi-distortionless networks, the saturation constraint, transient error minimization, and the design of servomechanisms. A radar engineer would probably select sections on finite-time filters and matched filters.

The Wiener solution provided the necessary mathematical tools for extensions of his work and for a unification of other earlier efforts. The extensions are all illustrations of the fact that if the calculus of variations works, it also works under constraints.

Section XV

North, Dwork, and Middleton independently derived the filters known variously as matched filters, comb filters, North filters, and Dwork filters. These men were interested maximizing the probability of detection of a pulse. Their solution was equivalent to minimizing the mean square noise output under the constraint that the filter duplicate the maximum value of the pulse. We will derive their results in a few lines with the benefit of a more powerful technique than was available to them.

Ragazzini and Zadeh re-derived the Dwork-North-Middleton results for a more general noise spectrum than flat noise and under the constraint that the network be realizable. Our derivation is a spectral density version of their time-domain solution.

Section XVI

Ragazzini and Zadeh also tried applying the constraint that polynomials in time pass through the network without distortion. This constraint is of necessity limited to finite-time filters. A filter which can remember the start of a polynomial must, by definition, contain the starting transient. Ragazzini and Zadeh require that the ensemble mean square error be minimized over the interval considered. This extension of Wiener's solution is straightforward but difficult of application. The design engineer usually has to solve five or six simultaneous equations to obtain the solution.

Section XVII

The Ragazzini and Zadeh, finite-time filter is usually approximated by the so-called quasi-distortionless network. This network passes polynomials without distortion except for the starting transient. The quasi-distortionless network can be optimized using either the Phillips or Wiener technique.

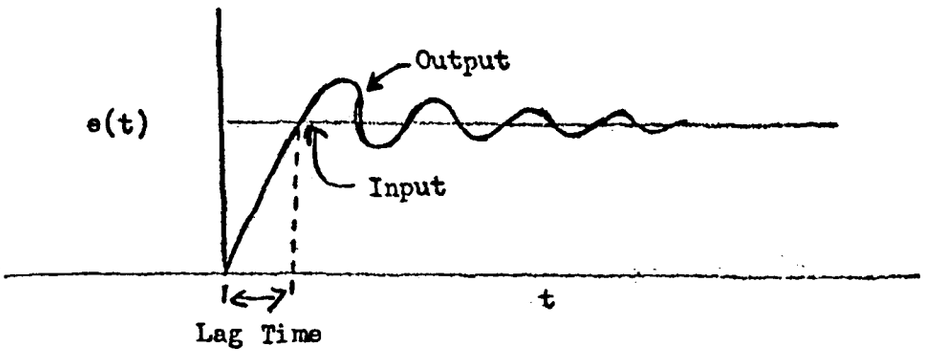
Section XVIII

Newton at M.I.T. and this author at J.P.L. considered systems in which saturation effects could so disrupt the system as to render operation worthless. Both solved the problem by applying a power level constraint at the appropriate system point.

Section XIX

The transient error of a system may be measured in many ways, depending upon the circuit application. One common definition is the integral of the transient-squared. Other definitions are certainly possible. In switching circuits the definition might be the lag time of the circuit -- the time

between reception of a step function and the first zero of the transient.



Both definitions are considered in deriving the best transient performance under noise conditions.

Section XX

R. J. Parks at J.P.L. applied Wiener's solution to guidance systems attempting to correct for external disturbances, and hindered by high level measurement noise. This development demonstrated an important point in philosophy: that f(t) is not necessarily a signal and n(t) not necessarily a noise in the expression for error. Both may be noise or both signals. The Wiener solution minimizes the "mean square error" defined by the error equation

$$\mathcal{E}(s) = H(s)\, N(s) + (1 - H(s))\, F(s)$$

where F(s) is earmarked not by the designation "signal" but by being the coefficient of 1-H(s). Similarly, N(s) is defined by being the coefficient of H(s). Any error equation of this form is treated by the Wiener solution.

Section XXI

The last section of the text discusses some related topics: multi-dimensional systems, short-time error systems, cross-correlation detection, and decision networks. The discussion is qualitative in nature. No solutions are presented.

XV MATCHED FILTERS

The matched-filter results presented here were partly known before, or independently of, Wiener. The results were derived originally using the Schwarz inequality. The basic problem was the detection of known pulse shapes under high noise conditions. The original statement of the problem asked that, at the output of the detection system, the peak signal strength be as large as possible relative to the output noise level. This criterion is equivalent to minimizing the mean output noise level under the constraint that the system duplicate the signal's peak value.

The signal is assumed zero until the (arbitrary) reference time $t = 0$. It then increases to its peak value (or any other value of interest) at a time t_o. The constraint we impose is that at t_o the system yield $f(t_o)$. The system will <u>not</u> yield $f(t)$ at other times unless by coincidence. The constraint is therefore given by:

$$f(t_o) = \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} e^{+st_o} H(s) F(s)\, ds \qquad (91)$$

$$= \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} \frac{e^{st_o} H(s) F(s) + e^{-st_o} H(-s) F(-s)}{2}\, ds$$

The problem to be solved is the minimization of the output mean square noise level under the constraint, i.e., the minimization of

$$\sigma^2 + \lambda f(t_o) = \frac{1}{2\pi j}\int_{-j\infty}^{j\infty}\left\{|H(s)|^2 \Phi_n + \frac{\lambda}{2}\left[e^{st_o}H(s)F(s) + e^{-st_o}F(-s)H(-s)\right]\right\} ds \tag{92}$$

Using the same plus and minus subscript notation and technique given in the Wiener derivation,

$$\sigma^2 + \lambda f(t_o) = \frac{1}{2\pi j}\int_{-j\infty}^{j\infty}\left\{H_+H_- \Phi_n + \frac{\lambda}{2}\left[e^{st_o}H_+F_+ + e^{-st_o}F_-H_-\right]\right\} ds \tag{93}$$

which is minimized by letting $H_+ \rightarrow H_+ + \varepsilon \eta_+$ and differentiating with respect to ε at $\varepsilon = 0$. The result is

$$0 = \int_{-j\infty}^{j\infty}\left\{\eta_-\left[H_+ \Phi_n + \frac{\lambda}{2} e^{-st_o} F_-\right] + \eta_+\left[H_- \Phi_n + \frac{\lambda}{2} e^{st_o} F_+\right]\right\} ds \tag{94}$$

By exactly the same technique as was used in the Wiener derivation, we obtain

$$H(s) = \frac{K}{\psi_n(s)}\left(\frac{e^{-st_o}F(-s)}{\psi_n(-s)}\right)_{\text{Physically Realizable}} \tag{95}$$

where $\psi_n(s)\psi_n(-s) = \Phi_n$ and where K is the gain constant necessary to yield the desired peak value.*

Dwork solved the problem without the restriction of physical realizability. His answer was $H(s) = Ke^{-st_o}F(-s)/\Phi_n$. This answer is physically-realizable under certain specialized conditions. To mechanize a filter, Dwork assumed $\Phi_{nn} = n^2$ and that f(t) had a finite time duration. As seen above, these expediencies are unnecessary. Under Dwork's conditions the H(s) becomes

---

*Note the caution on e^{-st_o} usage in XII D.

$$H(s) = Ke^{-st_o} F(-s) \qquad (96)$$

The matching of H(s) to F(-s) results in the name "matched filter" for this H(s). A similar result was derived in the section on curve fitting. The weighting function matched the form of the signal function.

These filters are of great interest in radar problems when the general form of the return echo (or chain of echoes) is known, but we wish the return signal to "stand out" of the noise as much as possible.

A more general result can be derived if the signal is assumed to have a statistical part as well as the known f(t). The answer is

$$H(s) = \frac{1}{\Psi(s)} \left(\frac{\Phi_{ff} + \Phi_{fn} + K F(-s)\, e^{-st_o}}{\Psi(-s)} \right)_{\substack{\text{Physically} \\ \text{Realizable}}} \qquad (97)$$

where

$$\Psi(s)\,\Psi(-s) = \Phi_{ff} + \Phi_{fn} + \Phi_{nf} + \Phi_{nn}$$

Problem 35: Prove the result above.

Problem 36: Show that the filter to maximize the ratio of the peak value of $f(t) = 1 \cdot e^{-\alpha t}$ in the presence of noise of spectrum $n^2/(1-s^2)$ is given by

$$H(s) = K \left(\frac{1+\alpha}{4n^2} \right) \left(\frac{1+s}{\alpha+s} \right)$$

where the approximation $e^{-x} = \frac{1}{1+x}$ is assumed.

The impulsive response and weighting function of a Dwork "matched filter" are of a particularly interesting form. From the Dwork expression for H(s) we find the impulsive response to be

$$h(t) = \frac{1}{2\pi j}\int_{-j\infty}^{j\infty} e^{st}F(-s)\,Ke^{-st_o}\,ds = \frac{K}{2\pi j}\int_{-j\infty}^{J\infty} F(s)e^{+s(t_o-t)}ds = Kf(t_o-t)$$

The impulsive response therefore starts with a value $f(t_o)$ and traces the signal backwards to its start.

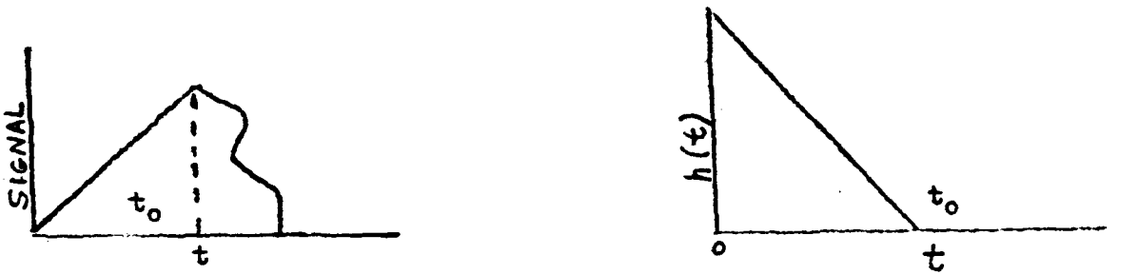


FIGURE 17

Matched Filter Impulsive Response

The weighting function dragged down the record thus looks like the first part of f(t) up to time t_o. For non-flat noise, this characteristic is no longer true.

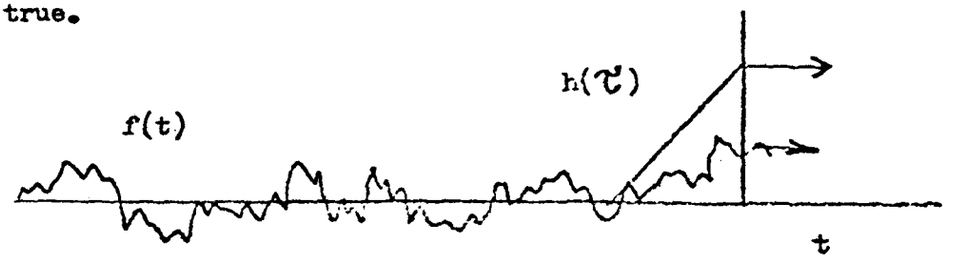


FIGURE 18

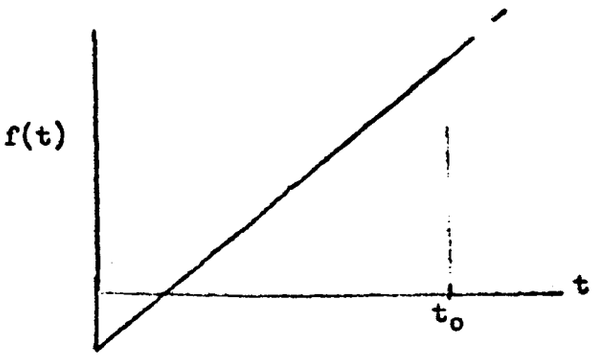
Matched Filter Superposition Integration

Problem 37: Sketch the Dwork weighting function for a code group of three equal-amplitude, equally-spaced pulses, assuming the event of greatest interest is the appearance of the last pulse after proper appearance of the first two pulses. Sketch the filter output to the code group as it appears. How might such a filter be realized using time-delays? Assuming that the center pulse is rectangular of width ΔT, show the effect of displacement of the center pulse from its expected position.

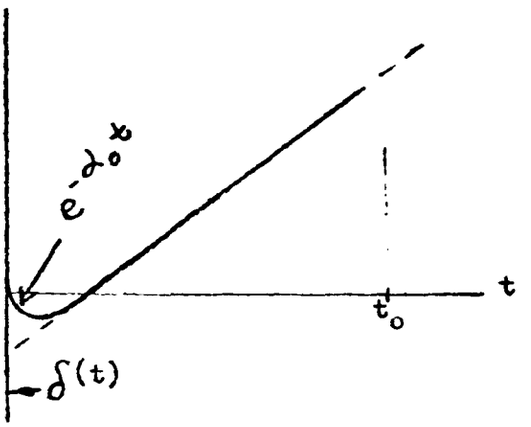
The weighting function of the more general filter given in equation (95) may be constructed from our knowledge of F(s) and $\Psi_n(s)$. In equation (95) (repeated below)

$$H(s) = \frac{1}{\Psi_n(s)} \left(\frac{F(-s)}{\Psi_n(-s)} e^{-st_o} \right)_{P.R.} \qquad (95)$$

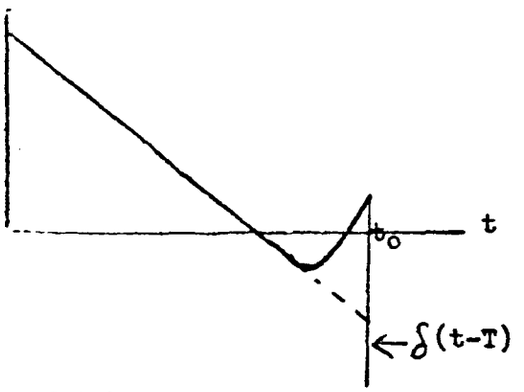
the impulsive response of H(s) is given by the impulsive response of $\left[F(-s)e^{-st_o}/\Psi_n(-s)\right]_{P.R.}$ passed through the network $1/\Psi_n(s)$. The impulsive response of the bracket, however, is the response of a network $1/\Psi_n(s)$ to f(t) played backwards from t_o. The graphical construction of the weighting function of H(s) is illustrated in the following sketches. A function, $f(t) = a + bt$, and a network $\Psi_n(s) = \frac{\alpha_o + s}{\beta_o + \beta_1 s + \beta_2 s^2}$, were assumed.



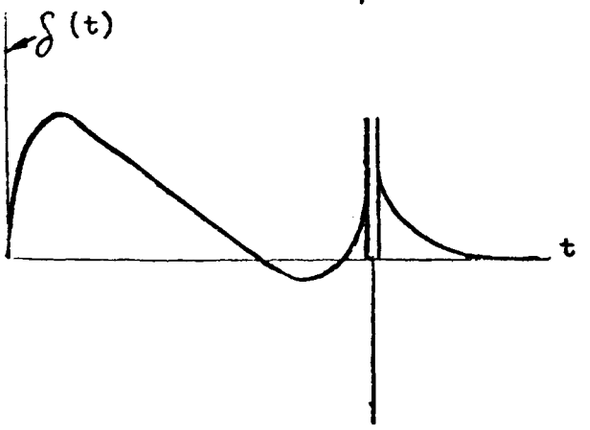
Response of $\overline{\psi_n'(s)}$ to f(t)



Impulsive Response of $\left(\frac{F(-s)}{\psi_n(-s)} e^{-st_o}\right)$ P.R.



Impulsive Response of H(s)



Needless to say, it is difficult to mechanize a network H(s) which will yield such an impulsive response. The mechanization of the unusual response near t_o would probably require delay lines. Rising exponentials are permitted in the interval $0<t\leq t_o$, but such exponentials are prohibited for t greater than t_o. The designer must therefore obtain exact cancellation of rising exponential terms after t_o.

The optimum network specified by equation (95) is primarily useful as a standard of comparison for proposed substitutes. The optimum realizable network yields

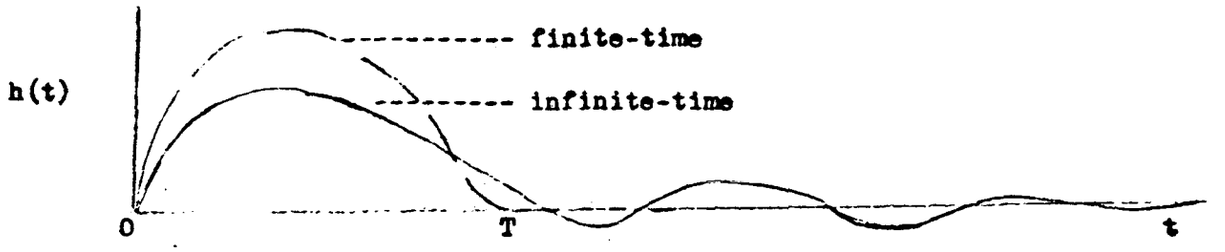
$$\frac{f^2(t_o)}{\sigma^2_{min.}} = \frac{\left[\frac{1}{2\pi j}\int_{-j\infty}^{j\infty}\left(\frac{F(-s)e^{-st_o}}{\Psi(-s)}\right)_{P.R.}\frac{F(s)e^{st_o}}{\Psi(s)}\,ds\right]^2}{\frac{1}{2\pi j}\int_{-j\infty}^{j\infty}\left|\left(\frac{F(-s)e^{-st_o}}{\Psi(-s)}\right)_{P.R.}\right|^2 ds}$$

The optimum non-realizable network yields

$$\frac{f^2(t_o)}{\sigma^2_{min.}} = \frac{1}{2\pi j}\int_{-j\infty}^{j\infty}\frac{|F(s)|^2}{\Phi_n}\,ds$$

XVI The Finite-Time, Finite-Order, Filter

A finite-time filter is one whose impulsive response lasts only for an interval T, after which it is identically zero. Most readily-mechanizable transfer functions are expressible as quotients of polynomials in s. The impulsive responses corresponding to such transfer functions are thus sums of exponentials in time. Consequently most readily-mechanizable networks do not exhibit finite-time behavior.



The optimization of finite-time systems is no different than the optimization of infinite-time systems, provided that our statistical averaging is done in an ensemble fashion (over many experiments). For stationary inputs, the spectral density approach is still legitimate; error spectral densities may be used as before. For the non-stationary or time-variant case, however, we must again distinguish between the mean square error averaged over an ensemble and the mean square error averaged over the interval T. This text will consider only the ensemble average case, although the interval average is more significant for the finite-time case than for the infinite-time case.

The meaning of the expression, finite-order, is closely connected with the application of constraints to optimization problems. In previous sections we have discussed constraints which specify response characteristics to a fairly general class of "known" signal inputs f(t). (Response at t_o, minimum lag time, etc.)

A still different set of constraints were proposed by Zadeh and Ragazzini in the first of two articles on the optimization of finite-time systems:*

$$\begin{aligned} \mu_o &= \int_0^{T_o} h(\tau)d\tau \\ \mu_1 &= \int_0^{T_o} \tau h(\tau)d\tau \\ &\vdots \\ \mu_n &= \int_0^{T} \tau^n h(\tau)d\tau \end{aligned} \qquad (98a)$$

These n constraints are used to guarantee distortion-free passage of a signal function

$$f(t) = a_o + a_1 t + a_2 t^2 + \dots a_n t^n$$

through the system inasmuch as it is easily shown that the response of the filter to f(t) is given by

$$e_{out}(t) = \mu_o f(t) - \mu_1 \dot{f}(t) + \frac{\mu_2}{2!} \ddot{f}(t) + \dots (-1)^n \frac{\mu_n}{n!} f^{(n)}(t)$$

Evidently if the best present value of f(t) is desired, $\mu_o = 1$, $\mu_1 = \mu_n = 0$. The name, finite-order, is applied

---

*1. Zadeh & Ragazzini: An Extension of Wiener's Theory, J.A.P. Vol.21, July 1950
2. Zadeh & Ragazzini: Optimum Filters for the Detection of Signals in Noise, I.R.E., October 1950.

to this filter because the polynomial is of finite-order, n. It might be remarked that any function can be inserted in constraint form on the system expressible as a sum of known functions of time such that the functional form does not change with a shift of the time origin. As an example:

$$f_1(t) = a \sin \beta t + b \cos \beta t \qquad (\beta \text{ known})$$

yields the constraints

$$\mu_a = \int_0^T \sin \beta\tau \, h(\tau) d\tau$$

$$\mu_b = \int_0^T \cos \beta(\tau) h(\tau) d\tau$$

The system output to $f_1(t)$ is

$$e_{out}(t) = \int_0^T f_1(t-\tau)\, h(\tau)\, d\tau$$

$$= \int_0^T a \sin(\beta t - \beta\tau)\, h(\tau) d\tau + \int_0^T b \cos(\beta t - \beta\tau) h(\tau) d\tau$$

$$= \mu_a \left[-a \cos\beta t + b \sin\beta t\right]$$

$$+ \mu_b \left[a \sin\beta t + b \cos\beta t\right]$$

Therefore if $\mu_b = 1$ and $\mu_a = 0$, the output is $f_1(t)$. If $\mu_b = 0$ and $\mu_a = -\beta$, the output is $f_1'(t)$. If both μ_a and μ_b are zero, $f_1(t)$ will not pass through the system.

The various problems considered so far -- the Wiener smoothing problem, the Wiener operational problem, and the matched filter problem -- could all be constrained by specifying $h(t > T) = 0$. The constraint specifying a finite-time filter is evidently

$$h(t>T) = 0 = \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} H(s) e^{sT} e^{st_1} ds \qquad (t_1 > 0)$$

It is easily shown that this constraint adds the term

$$\frac{\lambda}{\Psi(s)} \left(\frac{e^{-sT} e^{-st_1}}{\Psi(-s)} \right)_{P.R.}$$

to the H(s) already determined for the various problems considered. The $\Psi(s)$ is the appropriate one for the problem considered. For the Wiener case, $|\Psi(s)|^2 = \Phi_{ff} + \Phi_{fn} + \Phi_{nf} + \Phi_{nn}$; for the matched filter case $|\Psi(s)|^2 = \Phi_{nn}$

Inasmuch as not only h(t) but all of its derivatives and integrals (from T) are also zero for $t > T$, the most general set of constraints would result in the addition of a term

$$\frac{1}{\Psi(s)} \left(\frac{G(-s)\, e^{-sT} e^{-st_1}}{\Psi(-s)} \right)_{P.R.}$$

to the various H(s). The function G(s) must yet be determined; it represents a generalized Lagrange multiplier. Unfortunately, the variable t_1 is entangled in the answer; it is not obvious how to make the answer independent of t_1.

Nonetheless, several clues are provided:

1. The finite-time answer will consist of the infinite-time answer plus some additional terms.
2. The additional terms will probably be constants, terms of the form $AS + BS^2 + \ldots$, and terms arising from the roots of the numerator of $\Psi(s)$.
3. The weighting function will be modified both within the interval $0 < t < T$ and outside of it.

Ragazzini and Zadeh solved the problem in the time-domain. Their answers for $h(t)$ within the interval $0 < t < T$ are given below.

If the problem is to find the best finite-time filter for the Wiener case (where Φ_f and Φ_n are known and where the desired operation is $H_1(s)$) the answer is

$$h(t) = \mathcal{L}\left[\frac{1}{\Psi(s)}\left(\frac{H_1(s)\,(\Phi_{ff}+\Phi_{fn})}{\Psi(-s)}\right)_{P.R.}\right] \qquad (98b)$$

$$+ \sum_{j=1}^{K_1} A_j e^{\alpha_j t} + \sum_{j=1}^{K_2} B_j \delta^{(j)}(t)$$

$$+ \sum_{j=1}^{K_2} C_j \delta^{(j)}(t - T)$$

If the problem is the best matched filter, the finite-time answer is

$$h(t) = \mathcal{L}\left[\frac{K}{\psi^{n}(s)}\left(\frac{F(-s)e^{-st_o}}{\psi_r(-s)}\right)_{P.R.}\right] \qquad (98c)$$

$$+ \sum_{j=1}^{K_1} A_j e^{\alpha_j t} + \sum_{j=1}^{K_2} B_j \delta^{(j)}(t) + \sum_{j=1}^{K_3} C_j \delta^{(j)}(t-T)$$

where $T \geq t_o$.

In the particular case where the constraints are given in the form of equation 98a,

$$h(t) = \sum_{j=1}^{n} D_j t^n + \sum_{j=1}^{K_1} A_j e^{\alpha_j t} + \sum_{j=1}^{K_2} B_j \delta^{(j)}(t) \qquad (98d)$$

$$+ \sum_{j=1}^{K_2} C_j \delta^{(j)}(t - T)$$

The various solutions can be combined, inasmuch as the last two problems are solutions of the first problem under constraints. The summation terms are found as follows:

1. Write the $\psi(s)$ appropriate for the problem as a quotient of two polynomials:

$$\psi(s) = \frac{Q(s)}{R(s)} = \frac{q_o + q_1 s + \ldots . q_m s^m}{r_o + r_1 s + \ldots . r_p s^p}$$

2. Then $K_1 = m$

$$K_2 = p - m - 1$$

$$\alpha_j = \text{roots of } |Q(s)|^2 = 0$$

3. The coefficients A_j, B_j, C_j, and D_j must be found using the constraint equations and the appropriate optimization equation:*

Wiener: (Desired Operation is $h_1(\tau)$)

$$\int_0^T h(\tau)\,\varphi(t-\tau)d\tau = \int_{-\infty}^{\infty} h_1(\tau)\left[\varphi_{ff}(t-\tau) + \varphi_{fn}(t-\tau)\right]d\tau$$

Matched Filter:

$$\int_0^T h(\tau)\,\varphi\ (t-\tau)d\tau = f(t_o - t)$$

Finite Order:

$$\int_0^T h(\tau)\,\psi\ (t-\tau)d\tau = \lambda_o + \lambda_1 t + \ldots . \lambda_n t^n$$

(λ's must also be determined)

Ragazzini and Zadeh considered the Wiener-plus-finite-order filter in their first article and the matched filter in their second.

The finite-time filter is subject to the same difficulties of mechanization that plagued the matched filter. It is useful as a standard of comparison, nonetheless.

The finite-time filter is useful in solving the turn-on problem. The turn-on problem asks for the best filter such that the mean square error is minimized over the interval from t = 0 (turn-on) to the present time t = T. The turn-on filter is therefore a time-variant one depending on the parameter T.

The finite-time filter is also useful in data-reduction work in which the time interval is limited during which certain assumptions might be valid. For example, consider the problem of fitting polynomials to

---

*Readily derivable using the technique given in XII E. (See 85)

sections of a curve of complex shape. The polynomials are assumed valid only for short intervals. A finite-time weighting function of appropriate duration, T, will properly perform the fitting.

Problem 38a: What is the optimum finite-time filter to pass $a_2 + bt$ undistorted in the presence of noise $\Phi_n = \frac{2\alpha\sigma_n^2}{\alpha^2 - s^2}$?

Problem 38b: What is the optimum finite-time filter to find the derivative of a + bt, undistorted, in the presence of noise $\Phi_n = \frac{2\alpha\sigma_n^2}{\alpha^2 - s^2}$?

Problem 38c: Find the answers for parts a and b for a noise spectrum $\Phi_n = \frac{\alpha_o - s^2}{\beta_o - s^2}$. What is the answer if $\alpha_o = \beta_o$?

XVII QUASI-DISTORTIONLESS NETWORKS

Networks are quite common which produce no signal distortion in the steady state. A transfer function H(s) which satisfies H(0) = 1 will pass DC undistorted after the transient has died out. Similarly we may design networks which pass certain sine waves and reject others. One particular network has recently come into prominence because of its utility in short time duration problems. If a signal is so slowly varying that the experiment is completed long before a "cycle" might occur, an expansion of the signal in Fourier coefficients is almost meaningless. A much more legitimate expansion would be in powers of t. The particular network which passes polynomials in t undistorted in the steady state is called a quasi-distortionless network. Customarily the name, quasi-distortionless, is applied only to networks which leave the polynomial unaffected in the steady state. It is certainly possible to perform operations such as time delay or differentiation without steady-state distortion, but these networks are not commonly called quasi-distortionless. The quasi-distortionless network can perform other duties at the same time. Noise may be reduced; 60 cycle hum may be rejected.

The necessary condition for a network to be quasi-distortionless to a polynomial of order n is that its transfer function have matched numerator and denominator coefficients to s^n. The network transfer function in equation 99 is quasi-distortionless to t^2.

$$H(s) = \frac{a_0 + a_1 s + a_2 s^2 + a_3 s^3}{a_0 + a_1 s + a_2 s^2 + b_3 s^3 + b_4 s^4} \qquad (99)$$

The proof: In the steady state we may juggle the transfer function equation such that all s's represent differentiation.

$$e_{out} = H(s)\, e_{in}$$

$$(a_0 + a_1 s + a_2 s^2 + b_3 s^3 + b_4 s^4)e_{out} = (a_0 + a_1 s + a_2 s^2 + a_3 s^3)e_{in}$$

if $e_{out} = e_{in} = \alpha + \beta t + \gamma t^2$, the equation is satisfied identically.

Notice that the values of the a's and b's have not been specified. The choice of the a's and b's depends upon the further duties of the network and upon the allowable transient time of the system. The most useful extra duty of the quasi-distortionless network is noise reduction. For this purpose the quasi-distortionless network also must be optimized, although such an optimization is not very direct. Signals are rarely polynomials for more than a certain time interval. During this interval we wish the starting transient to die out as fast as possible. On the other hand, a quick transient implies a short-time weighting function and a wide-band system. The wide-band permits a great deal of noise to pass through the system. The engineer must find the best compromise.

Quasi-distortionless networks are actually optimum to a particular class of spectral densities. Remembering that a flat spectrum consists of a series of impulses, we can set up the following table.

SIGNAL SPECTRAL DENSITY	POSSIBLE TIME FUNCTION
1	impulses
$1/(s^2)$	steps
$1/s^4$	ramps
$(1/s^{2n})(-1)^n$	sections of polynomials of degree n
$1/(s^2)^N$	

The steps, ramps, etc. occur in a random manner as we have seen earlier. The noise spectral density apparently does not affect the fact that the H(s) specified for smoothing such signals are quasi-distortionless in form.* Inasmuch as the networks are determined by the Wiener operation, all coefficients are specified. No further engineering compromising is necessary.

Problem 30: Show that quasi-distortionless networks result as the optimum H(s) for the following signal and noise spectral densities.

Φ_{ff}	Φ_{NN}
$-1/s^2$	ε^2
$-1/s^2$	$\frac{\varepsilon^2}{1-s^2}$
$+1/s^4$	ε^2

The mean square signal values in all these cases are infinite due to the good probability (infinite DC energy) that the signal will drift to $\pm\infty$. Replacing s with $s+\alpha$ in the signal spectral densities will result in non-quasi-distortionless filters. Thus, the quasi-distortionless filter is a limiting case.

---

* Shown by example only.

XVIII THE SATURATION CONSTRAINT

Practical systems are linear only within certain power level limits. A system designed to handle 1 watt of power will usually saturate partially by 5 watts and totally by 10 watts.

The optimum system, as discussed to this point, has assumed perfect linearity of all components. Occasionally this assumption can lead to excessive power levels within the system. As an example, consider designing H(s) in the system below.

$$\Phi_{ff} = \left(\frac{1}{\alpha^2 - s^2} \right)$$

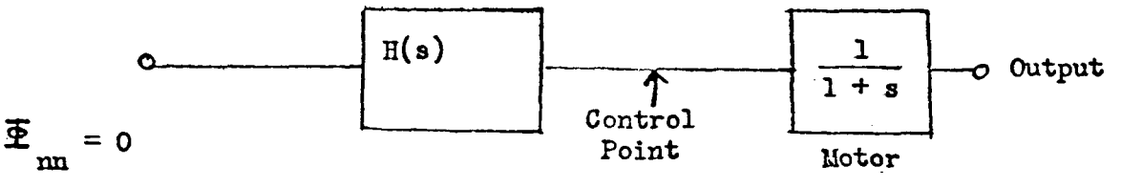


FIGURE 19

A Saturable System

In a perfectly linear system, H(s) = 1 + s. The input power level is equal to the output power level, neither of which is infinite. The power level at the control point, however, is infinite. Saturation of the network or motor is unavoidable.

The best solution to such a problem is the one which minimizes the output error but keeps the control point power, σ_c^2, within specified limits. The output error is given by

$$\sigma_\varepsilon^2 = \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} \left| 1 - H(s)\, H_m(s) \right|^2 \Phi_{ff}\, ds \qquad (100)$$

where H_M is the motor transfer function $1/(1+s)$. The power level at the control point is

$$\sigma_c^2 = \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} |H(s)|^2 \Phi_{ff} \, ds \qquad (101)$$

The quantity to be minimized is, therefore,

$$\sigma_\varepsilon^2 + \lambda \sigma_c^2 = \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} \left\{ |1-H(s)H_M(s)|^2 + \lambda |H(s)|^2 \right\} \Phi_{ff} \, ds \qquad (102)$$

The minimization might be done by the standard technique. A useful shortcut may be used in cases similar to this, however. The shortcut consists of re-writing the equation of interest (102) so that it resembles an equation that has already been solved. In this particular case we note the general resemblance of equation 102 to the equation describing the mean square error in the standard Wiener smoothing problem:

$$\sigma^2 = \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} \left\{ |1 - H(s)|^2 \Phi_{ff} + |H(s)|^2 \Phi_{nn} \right\} ds$$

In this standard smoothing equation, Φ_{ff} and Φ_{nn} are completely specified. H(s) is to be found. A substitution $W(s) = H(s) H_M(s)$ yields

$$\sigma_\varepsilon^2 + \lambda \sigma_c^2 = \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} \left\{ |1 - W(s)|^2 \Phi_{ff} + \lambda |W(s)|^2 \frac{\Phi_{ff}}{|H_M(s)|^2} \right\} ds \qquad (103)$$

This problem has already been solved for $W(s)$ if $\Phi_{ff} / |H_M(s)|^2$ is thought

of as an "equivalent noise spectrum". Using this shortcut, it is straightforward to show that

$$H(s) = \frac{1}{\Psi_F(s)\Psi_M(s)} \left(\frac{H_M(-s)\Psi_F(s)}{\Psi_M(-s)} \right)_{\substack{\text{Physically} \\ \text{Realizable}}} \qquad (104)$$

where $\Psi_F(s)\,\Psi_F(-s) = \Phi_{ff}$ and where $\Psi_M(s)\,\Psi_M(-s) = (\lambda + |H_M(s)|^2)$

In the problem illustrating control of a particular motor:

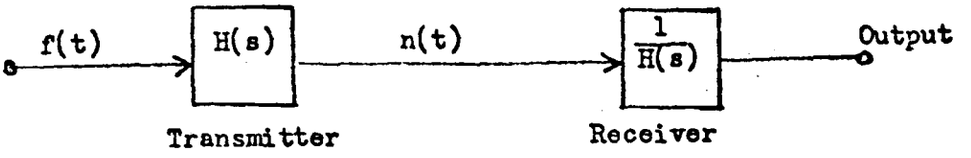
$$H(s) = (1+s)\left[\frac{1}{\alpha\sqrt{\lambda} + \sqrt{\lambda+1}} \cdot \frac{1}{\sqrt{\lambda+1} + \sqrt{\lambda}\,s} \right]$$

If the power constraint were not present ($\lambda = 0$) we would obtain the expected answer $H(s) = 1 + s$. If the power constraint is very severe (large values of λ) then $H(s) \longrightarrow \frac{1}{\lambda(1+\alpha)}$, describing an attenuator. The value of λ is chosen to yield the specified σ_c^2.

The saturation constraint can usually be applied and solved in the shortcut manner illustrated. The equation to be minimized is re-written in the standard smoothing form. New equivalent signal and noise spectral densities are defined. The smoothing answer is written immediately and used to find the answer to the original problem.

Problem 40: Solve the standard smoothing problem under the constraint that the output noise power be limited to a specified value. Notice that this problem is equivalent to weighting noise error and signal error differently.

Problem 44: Solve for the H(s) giving the minimum output noise level in the transmission system below. Assume that $\Phi_f = \frac{1}{\alpha^2 - s^2}$ and $\Phi_n = \frac{1}{\beta^2 - s^2}$ and that the transmitter power is limited to a fixed value, P_T.



XIX TRANSIENT ERROR MINIMIZATION

The optimization technique may be applied to other problems than the minimization of mean square error. Engineers designing control systems are usually interested in a design giving the minimum transients to certain specified input signals. The definitions of minimum transient are varied. One common specification asks that the integral of the square of the transient be as small as possible.* If $\mathcal{E}(t)$ is the transient, f(t) the known input, and H(s) the system transfer function, then

$$\text{Transient Error} \equiv \int_0^\infty \mathcal{E}^2(t)dt = \frac{1}{2\pi j}\int_{-j\infty}^{j\infty} |1-H(s)|^2 |F(s)|^2 \, ds \quad (105)$$

If the minimization of this integral were the only requirement there would be no problem. H(s) = 1 satisfies this condition. Typically, the constraints are those of proper performance in the presence of noise or operation within a specified power level.

---

* The definition of transient error is almost like the definition of signal distortion in the spectral density case, except that in the latter case, we take a time average.

$$\text{Transient Error} \equiv \int_0^\infty \mathcal{E}^2(t)dt$$

$$\text{Signal Distortion in Spectral Density Problems} = \lim_{T\to\infty} \frac{1}{T}\int_0^T \mathcal{E}^2(t)\,dt$$

For example, consider minimizing the transient error of the system of figure 19 under the condition that the output noise level be less than σ_o^2.

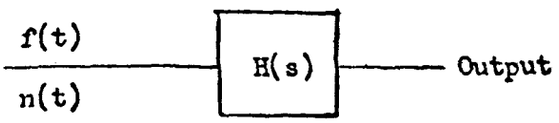


FIGURE 19

Illustrative System

The integral to be minimized is therefore

$$\int_0^\infty \varepsilon^2(t)dt + \lambda\sigma_o^2 = \frac{1}{2\pi j}\int_{-j\infty}^{j\infty}\left\{|1-H(s)|^2|F(s)|^2 + \lambda\Phi_n|H(s)|^2\right\}ds \qquad (106)$$

The equation is almost an exact copy of the Wiener smoothing problem equation. The answer is therefore

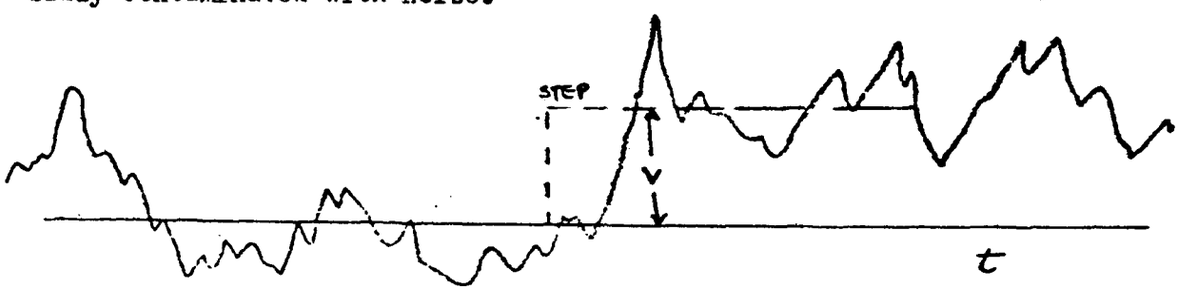
$$H(s) = \frac{1}{\Psi(s)}\left(\frac{|F(s)|^2}{\Psi(s)}\right)_{\text{Physically Realizable}} \qquad (107)$$

where $|\Psi(s)|^2 = |F(s)|^2 + \lambda\Phi_n$ and where λ is so chosen that the output noise level be less than σ_o^2. Notice from equations 106 and 107 that the solution is the same as that for a signal spectral density $\Phi_{ff} = n|F(s)|^2$ where $n = 1$. Therefore the Wiener solution minimizes the transient error to a single occurrence of $\Psi_{ff}(s)$ where $|\Psi_f(s)|^2 = \Phi_{ff}(s)$.

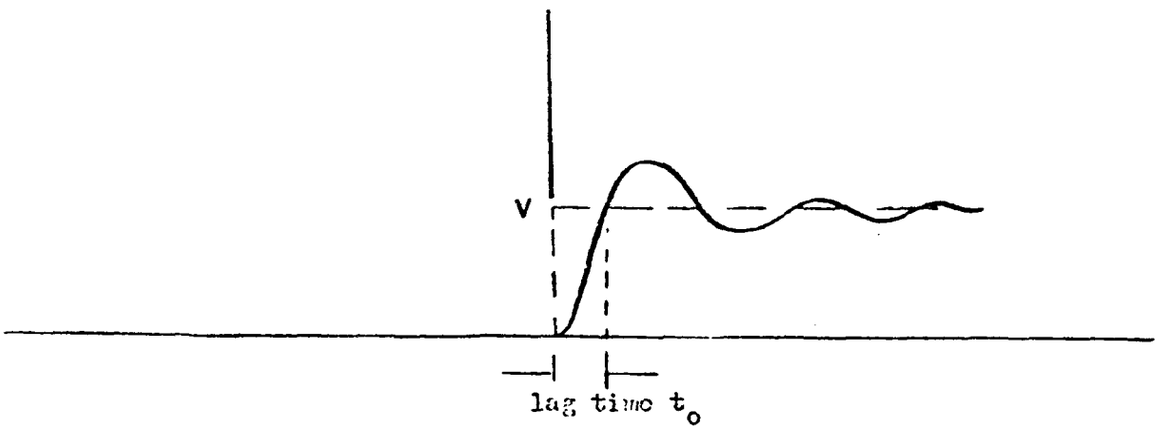
Problem 42: Minimize the mean square noise output of a system under the constraint that the transient error due to a step input be less than some specified constant E. Assume $\Phi_n = n^2$.

Problem 43: Minimize the mean square noise output of a system under the double constraint that the transient error to a step be E_1 and to a ramp, E_2. Assume $\Phi_n = n^2$.

The definition chosen for transient error depends upon the system. For example, consider the problem of quick-closing a switch with a step input badly contaminated with noise.



Assume that the switch closes and locks upon reception of a voltage of level V. The criterion for optimization is the minimization of the lag time of the circuit under the constraint that the output noise level be less than $(V/10)^2$. The lag time is defined as the time interval between the start of a (noise-free) step and the instant that the circuit output first reaches V.



The time that the circuit reaches V is given by

$$V = \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} \left[H(s)\right] \frac{V}{s} \, e^{st_o} \, ds \qquad (108)$$

The problem is considerably simplified by use of the symmetry condition of the calculus of variations. Instead of minimizing t_o with the output noise level fixed, minimize the output noise level with t_o fixed. Fixing t_o, however, is equivalent to the constraint $\mathcal{E}(t_o) = 0$. Therefore the integral to be minimized is

$$\sigma^2_{out} + \lambda \mathcal{E}(t_o) = \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} \left\{ \left|H(s)\right|^2 \Phi_n \right. \, ds$$

$$+ \frac{\lambda}{2} \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} \left\{ \left[1-H(s)\right] \frac{V}{s} \, e^{st_o} + \left[1-H(-s)\right] \frac{V}{(-s)} \, e^{-st_o} \right\} ds$$

The constraint is written as functions of H(s) and H(-s) for convenience. The integral equation, however, is exactly the same as that used in the matched filter derivation. The answer is therefore

$$H(s) = \frac{K}{\Psi_n(s)} \left(\frac{V/s \, e^{-st_o}}{\Psi_n(-s)} \right)_{P.R.} \qquad (109)$$

where K is a gain constant to be determined, $\left|\Psi_n(s)\right|^2 = \Phi_n$ and where t_o is the lag time. The solution is completed by finding the relationship between the output noise level and t_o. If $\Phi_n = n^2$, the best h(t) is a constant of

height $1/t_o$ lasting for t_o seconds. The output noise level is $\sigma_o^2 = n^2/t_o$. The output to a step input is a ramp function, 0 at $t = 0$ and V at t_o. After t_o, the output is a constant, V. Setting $\sigma_o^2 = (V/10)^2$ as originally specified, $t_o = 100n^2/V^2$ seconds.

XX THE DESIGN OF SERVOMECHANISMS

Two major techniques have been presented which may be used to design servomechanisms operating in the presence of interference. The first is the Phillips technique. The mean square error of the system is expressed as a function of the input spectral density constants and of the adjustable parameters of the system. The mean square error is then minimized by proper adjustment of the parameters. The complete form of the system is specified.

The second major technique stems from Wiener's optimization solution. The system is now assumed known except for a transfer function H(s). It is the goal of the technique to specify H(s) completely. In a servomechanism, such a specification determines the loop transfer function, the loop gain, and the loop transient behavior. For example, consider the simple servo illustrated below.

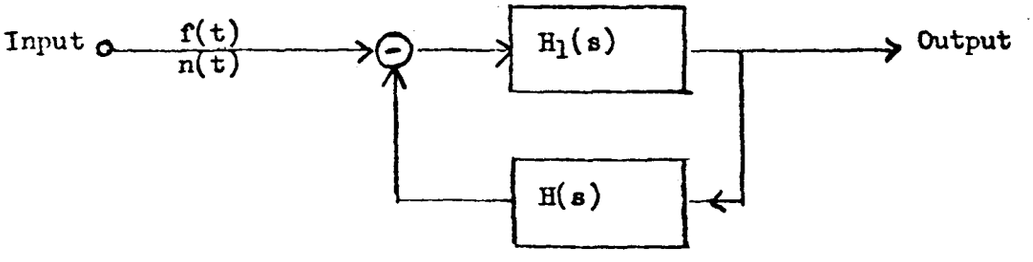
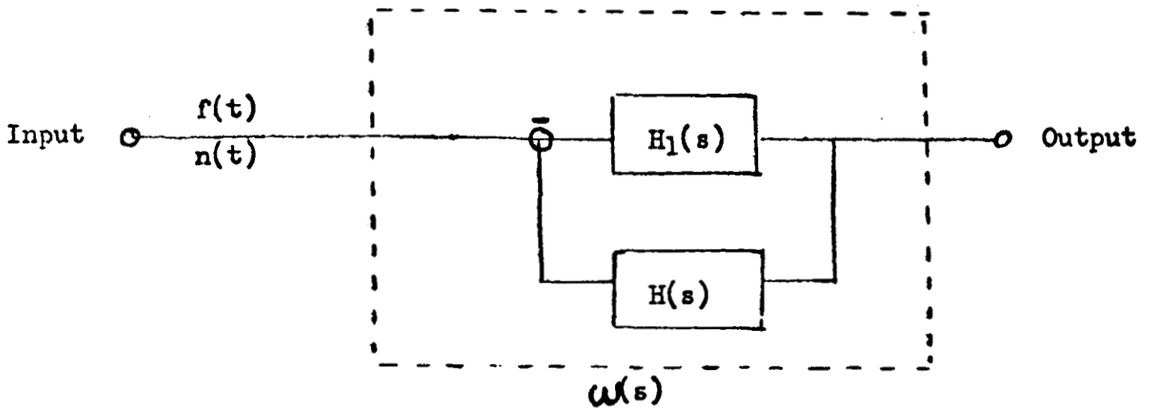


FIGURE 20

A Simple Servomechanism

Given $H_1(s)$, what is the best design for H(s) to yield the minimum mean square following error? The solution here is most simply achieved by replacing the whole servo with $\overline{W}(s)$.



$$\omega(s)_{Opt.} = \frac{1}{\psi(s)}\left(\frac{\Phi_{ff}+\Phi_{fn}}{\psi(-s)}\right)_{\substack{\text{Physically}\\\text{Realizable}}} = \frac{H_1(s)}{1+H_1(s)H(s)} \qquad (110)$$

Knowing $\omega(s)$ and $H_1(s)$ we solve for $H(s)$ to complete the problem. The important feature of this technique is the complete specification of $H(s)$ and consequently of the whole loop.

Servomechanisms designed in this manner quite often run afoul of power limitations. As we have seen*, however, power level constraints may be added without difficulty and the problem usually solved by defining equivalent noise and signal spectral densities.

If the servo is asked to perform an operation, $H_2(s)$, rather than simply following the input, the best design is evidently given by

$$\omega(s)_{Optimum} = \frac{1}{\psi(s)}\left(\frac{(\Phi_{ff}+\Phi_{fn})H_2(s)}{\psi(-s)}\right)_{\substack{\text{Physically}\\\text{Realizable}}} = \frac{H_1(s)}{1+H_1(s)H(s)} \qquad (111)$$

*Section XVIII

The foregoing illustrative servo is a quiet servo in that no noise or disturbance is introduced within the loop itself. The problem of the noisy servo can also be solved, however. This solution was first demonstrated to this author's knowledge by R. J. Parks of J.P.L. Consider the following noisy servo.

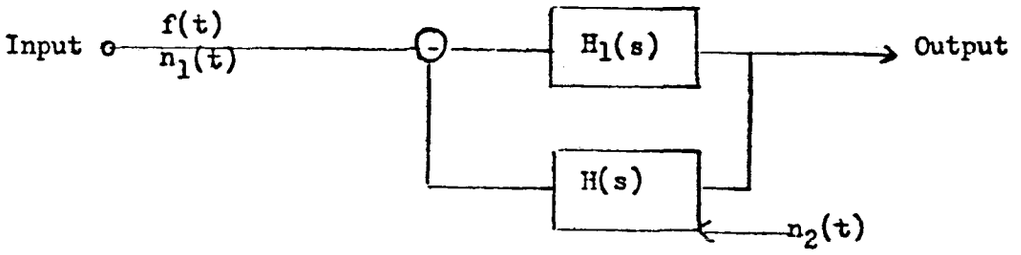


FIGURE 21

A Noisy Servomechanism

The second noise, $n_2(t)$, might be measurement error of the output such as potentiometer noise from an output position indicator. If the servo is connected together by radio links, $n_2(t)$ might be atmospheric static. For illustration, assume the servo is trying to follow f(t) and that all signals and noises are independent of each other. The error of such a system is:

$$E(s) = \left[\frac{H_1}{1 + H_1 H} n_1 + \frac{H_1}{1 + H_1 H} F - \frac{H H_1}{1 + H_1 H} n_2 \right] - F \qquad (112)$$

The problem will be solved by reducing it to an equivalent smoothing problem*

* For a simpler illustration of this method of solution see section XVIII. Reducing a problem to the equivalent standard smoothing problem lets us use the standard smoothing answer. The direct method involves another complete derivation.

This particular reduction is especially interesting because it illustrates the use of cross-spectral densities in a solution even though all the original inputs are assumed uncorrelated.

Reviewing the quiet servo, notice that the solution was simplified by using $\mathcal{W}(s)$ instead of H(s) before starting the operations of factoring, etc. The noisy servo solution is quite similar. The substitution is slightly different.

Let $$\mathcal{W}(s) = \frac{H(s)H_1(s)}{1 + H(s)H_1(s)} \qquad (113)$$

Then
$$E(s) = (1-\mathcal{W})\, n_1H_1 + (1-\mathcal{W})H_1F - \mathcal{W}n_2 - F \qquad (114)$$

$$= (1-\mathcal{W})\left[H_1n_1 + H_1F-F\right] - \mathcal{W}\left[n_2+F\right]$$

The equivalent signal and equivalent noise are defined by analogy to the simple smoothing error equation:

$$E(s) = (1-\mathcal{W})\, f'(s) + \mathcal{W}n'(s) \qquad (115)$$

Thus, the equivalent signal and noise are

$$\text{Equivalent Signal*} = H_1n_1 + H_1F-F \qquad (116)$$

$$\text{Equivalent Noise*} = -(n_2 + F)$$

Notice the cross-correlation between equivalent signal and equivalent noise due to the presence of F in both functions.

---

* Neither the equivalent signal nor the equivalent noise may be functions of the unknown H(s) by definition of the basic Wiener problem.

Writing the equivalent (primed) spectral densities:

$$\begin{aligned}
\Phi_{f'f'} &= \Phi_{n_1n_1}|H_1|^2 + |H_1 - 1|^2\Phi_{ff} \\
\Phi_{n'n'} &= \Phi_{n_2n_2} + \Phi_{ff} \\
\Phi_{n'f'} &= -\left(H(-s) - 1\right)\Phi_{FF} \\
\Phi_{f'n'} &= -\left(H(s) - 1\right)\Phi_{FF}
\end{aligned} \tag{117}$$

The solution may be written by analogy to the simple smoothing problem answer.

$$\omega(s) = \frac{1}{\psi'(s)} \left(\frac{\Phi_{f'f'} + \Phi_{f'n'}}{\psi'(-s)}\right)_{\substack{\text{Physically} \\ \text{Realizable}}} \tag{118}$$

where

$$\psi'(s)\,\psi'(-s) = \Phi_{f'f'} + \Phi_{f'n'} + \Phi_{n'f'} + \Phi_{n'n'} \tag{119}$$

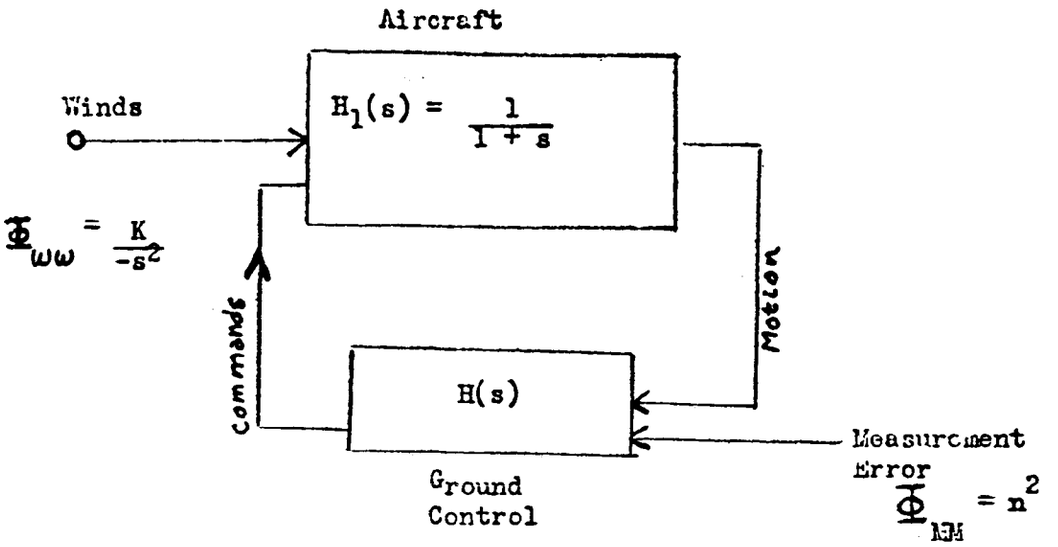
and where

$$\omega(s) = \frac{H_1(s)\,H(s)}{1 + H_1(s)\,H(s)} \tag{120}$$

An interesting point is driven home if we allow f(t) to be zero. No formal signal enters the network, only disturbances. The error of the system is still expressible in the equivalent smoothing form. The equivalent signal is n_1H_1 and the equivalent noise n_2. No cross-spectral densities exist.

THUS, ANY ERROR EXPRESSION RESEMBLING THE ERROR EXPRESSION IN THE BASIC WIENER PROBLEM MAY BE MINIMIZED BY THE WIENER OPTIMIZATION EQUATION.

Problem 44: Design the system below to best counteract wind disturbances under severe measurement error conditions.



XXI RELATED TOPICS

This text has endeavored to present certain fundamentals of linear system design. At least four related subjects have been almost totally ignored.

A. Multi-dimensional Systems

The systems discussed have all been one-dimensional. A multi-dimensional system would be one where the best operations on many different (but related) phenomena were to be specified.

B. Short-time Error Systems

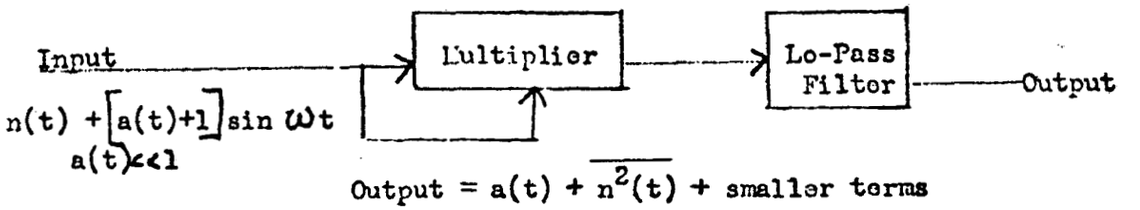
In certain cases in which the total time of operation is short compared to the time necessary to reach steady state, the long-time average of σ_E^2 may be unimportant. It may even be profitable to design an unstable system. As an example, consider the integration of a function in the presence of flat noise. With the system at rest until $t = 0$, the output error at $t = 0$ will be zero. At $t = \infty$, however, the mean square error is infinite. At intermediate times, the expected error may still be acceptable.

C. Cross-correlation Detection

If the exact time dependence of a signal is known except for a few parameters, the parameters may be determined by correlating the total input with a local signal or set of signals. For example, a sine wave may be detected in the presence of noise by cross-correlating the total input with a local sine wave in phase with the incoming sine wave. The technique is widely used under the names product detection and cross-correlation detection. Such detection systems are inherently better

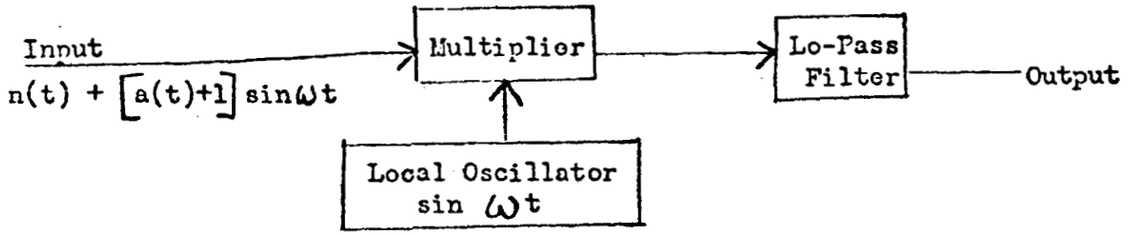
than auto-correlation detectors but are more complex in requiring a stable local source.

AUTO-CORRELATION DETECTION



$$\text{Output} = a(t) + \overline{n^2(t)} + \text{smaller terms}$$

CROSS-CORRELATION DETECTION



$$\text{Output} = a(t) + \text{smaller terms}$$

FIGURE 22

A Comparison of Auto-Correlation Detection with Cross-Correlation Detection

D. Decision Networks

The cross-correlation detector is a good example of a system using considerably more information than provided by spectral densities. In Figure 22, for example, the exact frequency and phase of the modulation sine wave were known.

Decision networks provide another example of a high degree of a priori knowledge. If little but the spectral density of a signal is known, the

system must be able to produce an infinity of possible output values. On the other hand, if it is known beforehand that the signal is either on or off, and that it consists of a pure sine wave of known amplitude and phase, the problem is quite different. Networks which decide the question of the presence of functions are examples of the general class, "decision networks." Such networks are often non-linear. The decision is often indicated by the opening or closing of a switch. Decision networks are designed using the statistical techniques of Neyman-Pearson, Siegert, and Wald.

REFERENCE LIST

I. The Criteria for Optimum

A. Bode & Shannon: Linear Least Square Smoothing and Prediction Theory, April 1950 IRE, pages 424 and 425: "Discussion of the Basic Assumption."

B. Wiener: Stationary Time Series, page 13

C. Lawson and Uhlenbeck: Threshhold Signals, pages 149-151, 161-167.

D. Golay: Correlation vs. Linear Transforms, February 1953 IRE, pages 268 and 269.

II. Elementary Calculus of Variations

A. Houston: Principles of Mathematical Physics, Chapter V, pages 53 - 61.

B. Wiener: Stationary Time Series, page 14

III. Constraints

A. Houston: Principles of Mathematical Physics, Chapter V, pages 53 - 61

B. Phillips: Analytical Geometry and Calculus, Chapter XVII, Section 209

IV. The Curve Fitting Problem and Weighting Functions

A. Kenney: Mathematics of Statistics, pages 145 - 150, (The Classical treatment) pages 152-153, (relative to problem 6: this fit is not the best if the noise is on y.)

B. Hoel: Introduction to Mathematical Statistics, Chapter V, pages 78 - 92.

BIBLIOGRAPHY

V FOURIER AND LAPLACE TRANSFORMS

A. Gardner and Barnes: Transients in Linear Systems, Chapters III and IV

B. Wiener: Stationary Time Series, pages 25-31, 36-37, 53-55

C. James, Nichols, & Phillips: Theory of Servomechanisms, Chapter 2, pages 23-62

VI THE SUPERPOSITION INTEGRAL

A. James, Nichols, & Phillips: Theory of Servomechanisms, Chapter 2, pages 30-40

B. Gardner and Barnes: Transients in Linear Systems, pages 228-236

VII CORRELATION FUNCTIONS

A. James, Nichols, and Phillips: Theory of Servomechanisms, pages 270-278. Note: Subscript convention used by these authors is the reverse of that used in this text.

VIII SPECTRAL DENSITY

A. James, Nichols, andPhillips: Theory of Servomechanisms, pages 278-291

B. Lawson and Uhlenbeck: Threshold Signals, Chapters 3 and 4. Note: These authors use a spectral density defined over positive frequencies only. This usage results in the expression for a particular spectral density being twice that used in this text.

IX THE FLAT NOISE CONCEPT

A. James, Nichols, and Phillips, Theory of Servomechanisms, pages 298-300, 321-322.

B. Lawson and Uhlenbeck: Threshold Signals, pages 42-46, 69-71.

X THE ERROR SPECTRAL DENSITY

A. James, Nichols, and Phillips: Theory of Servomechanisms, pages 312-315, 317-321

XI PHILLIPS OPTIMIZATION TECHNIQUE

A. James, Nichols, and Phillips: Theory of Servomechanisms, pages 308-315, 323-328, 369-370.

XII WIENER OPTIMUM FILTERS

A. Wiener: Stationary Time Series, pages 81-94, 129-133, 149-160

B. Richard C. Booton, Jr.: An Optimization Theory for Time-Varying Linear Systems, Proc. IRE, Vol. 40, pages 977-981, August 1952

XIII A PHYSICAL INTERPRETATION OF THE WIENER OPTIMUM

A. Bode and Shannon: A Simplified Derivation of Linear Least Square Smoothing and Prediction Theory, Proceedings of the IRE, April 1950, pages 417-425

XIV EXTENSIONS OF WIENER'S SOLUTION

XV MATCHED FILTERS

A. Dwork: Detection of a Pulse Superimposed on Fluctuation Noise, Proceedings of the IRE, Vol. 38, pages 771-774, July 1950

B. Zadeh and Ragazzini: Optimum Filters for the Detection of Signals in Noise, Proceedings of the IRE, Vol. 40, References given to filters on pages 1223, 1224, October 1952

XVI THE FINITE-TIME, FINITE-ORDER, FILTER

A. Zadeh, and Ragazzini: An Extension of Wiener's Theory, J.A.P., Vol. 21, July 1950, pages 645-655.

B. Zadeh and Ragazzini: Optimum Filters for the Detection of Signals in Noise, Proceedings IRE, Vol. 40, pages 1223-1231

XVII QUASI-DISTORTIONLESS NETWORKS

None readily available

XVIII THE SATURATION CONSTRAINT

A. George C. Newton, Jr.,: Compensation of Feedback-Control Systems Subject to Saturation, Journal of the Franklin Institute, Vol. 254, No. 4, October 1952, pages 281-413

XIX TRANSIENT ERROR MINIMIZATION

See reference in XVIII

XX THE DESIGN OF SERVOMECHANISMS

See reference in XVIII